

یک الگوریتم ممتیک ترکیبی جدید مبتنی بر PSO-SVM

برای انتخاب ویژگی

احسان اسلامی^۱، مهدی افتخاری^۲

^۱ دانشجوی کارشناسی ارشد، مهندسی کامپیوتر- هوش مصنوعی، بخش مهندسی کامپیوتر دانشگاه شهید باهنر، کرمان،

ehsan.eslami@eng.uk.ac.ir

^۲ استادیار، بخش مهندسی کامپیوتر، دانشگاه شهید باهنر، کرمان،

m.eftekhary@uk.ac.ir

چکیده

با پیشرفت سریع تکنولوژی کامپیوتری، پایگاه داده‌هایی با صدها و هزاران ویژگی در زمینه‌های شناسایی الگو، داده‌کاوی، یادگیری ماشین و غیره به وجود آمده است. به طوری که پردازش مجموعه داده‌های بزرگ، یک کار چالش‌برانگیز شده است، بنابراین با انتخاب بهترین مجموعه ویژگی، می‌توان به دقت قابل قبولی در پردازش پایگاه داده و همچنین در کاوش روابط معنادار بین ویژگی‌ها رسید. در این مقاله یک مدل ترکیبی جدید ارائه شده است که الگوریتم ازدحام ذرات^۱ را با الگوریتم ماشین بردار پشتیبان^۲ برای بهبود دقت طبقه‌بندی و انتخاب مجموعه ویژگی بهینه ترکیب می‌کند. این مکانیزم بهینه‌سازی، با ترکیب الگوریتم ازدحام ذرات گسسته و الگوریتم ازدحام ذرات پیوسته به طور همزمان زیرمجموعه ویژگی‌های بهینه را انتخاب و پارامترهای ماشین بردار پشتیبان را تنظیم می‌کند. در این الگوریتم از دو روش جستجوی محلی برای تنظیم پارامترها و بهبود زیرمجموعه ویژگی استفاده شده است. عملکرد روش پیشنهادی بر روی مجموعه داده‌های مختلف ارزیابی شد و نتایج حاصله نشان‌دهنده کارایی روش مزبور در داده‌های با ابعاد معمولی و ابعاد بالا در مقایسه با نسخه‌های دیگر می‌باشد.

کلمات کلیدی

انتخاب ویژگی، الگوریتم‌های ممتیک، جستجوی محلی، الگوریتم ازدحام ذرات، ماشین بردار پشتیبان

دقت بالاتری است. روش ترکیبی در تلاش است که از نقاط قوت هر دو روش فیلتر و فراگیر بهره‌برداری کند.

الگوریتم‌های ممتیک، یک فرایند جستجوی محلی را با یک فرایند جستجوی سراسری ترکیب می‌کند که الگوریتم جستجوی سراسری باعث قدرت کاوش الگوریتم کلی می‌شود و عملگرهای جستجوی محلی باعث افزایش قدرت الگوریتم در یافتن بهینه‌های محلی می‌شود که در این صورت قدرت بهره‌وری الگوریتم افزایش می‌یابد.

در این مقاله یک روش ترکیبی جدید انتخاب ویژگی و تنظیم پارامتر مبتنی بر الگوریتم PSO و SVM به همراه روش‌های جستجوی محلی ارائه شده است که به طور همزمان زیرمجموعه ویژگی بهینه را انتخاب و پارامترهای SVM را تنظیم می‌کند. در مدل SVM عواملی که باعث بهبود دقت طبقه‌بندی می‌شوند عبارتند از: ۱- محدود کردن تعداد ویژگی‌ها در طبقه‌بندی برای تولید مدلی با پیچیدگی محاسباتی کمتر و پیش‌بینی دقیق‌تر ۲- تنظیم

۱- مقدمه

انتخاب ویژگی، فرایند انتخاب بهترین ویژگی از میان تمام ویژگی‌ها است، فرایند انتخاب ویژگی این مساله را با از بین بردن داده‌های بی-رابط، زائد یا نویزی حل می‌کند. این عمل کارایی الگوریتم یادگیری را بهبود می‌بخشد، هزینه‌های محاسباتی را کاهش می‌دهد و درک بهتری را از مجموعه داده‌ها فراهم می‌کند.

روش‌های انتخاب ویژگی به سه روش فیلتر، فراگیر^۳ و ترکیبی^۴ تقسیم می‌شوند، روش فیلتر به خصوصیات و روابط موجود در پایگاه داده بستگی دارد و روند انتخاب ویژگی بدون هیچ مدل آموزشی انجام می‌شود. اما روش فراگیر نیاز به یک مدل از پیش تعیین شده دارد که هدف آن انتخاب ویژگی همراه با بهبود قابلیت تعمیم و کارایی است، روش فراگیر از لحاظ زمانی پرهزینه‌تر از فیلتر بوده، اما معمولاً دارای