Journal of Applied And Basic Machine Intelligence Research (ABMIR)

پژوهش‌های نظری و کاربردی هوش ماشینی

سال اول، شماره اول، بهار و تابستان ۱۴۰۲، صفحات ۵۹-۴۶

ارائه نسخه الگوریتم سینوس کسینوس چندگانه در حل مسئله انتخاب ویژگی

فاطمه سعادت جو[1*]، سهیل اقبالی[2]، علیرضا پورسلیمان[2]

[1] استادیار گروه مهندسی کامپیوتر، دانشگاه علم و هنر، یزد، ایران

[2] دانشجوی کارشناسی ارشد گروه مهندسی کامپیوتر، دانشگاه علم و هنر، یزد، ایران

چکیده **مقاله پژوهشی**

ازآنجاکه تمام ویژگی‌های داده‌ها برای یافتن دانشی که در آن‌ها نهفته است مهم و حیاتی نیستند؛ کاهش ابعاد ویژگی آن‌ها یکی از مباحث بااهمیت است. ازاین‌رو در این مقاله روشی جدید با استفاده از الگوریتم سینوس کسینوس با رویکرد بهینه‌سازی چندگانه در حوزه انتخاب ویژگی ارائه می‌شود. روش پیشنهادی در مدل انتخاب ویژگی رپر (مجموعه الگوریتم‌های تابع ارزیابی مبنی بر نرخ خطای طبقه بندی کننده) ارائه‌شده است و دو مرحله دارد که شامل مرحله انتخاب ویژگی با استفاده از الگوریتم سینوس کسینوس چندگانه و مرحله طبقه‌بندی جواب‌های ممکن در الگوریتم سینوس کسینوس با روش نزدیک‌ترین همسایه توسعه‌یافته است. روش پیشنهادی بر روی مجموعه داده استاندارد UCI در مجموعه داده‌هایی با ابعاد مختلف آزمایش‌شده است. مقایسه روش پیشنهادی با روش‌های بهینه‌سازی چندگانه و تک گانه نشان می‌دهد به‌طورکلی روش پیشنهادی در کاهش تعداد ویژگی‌ها توانسته نسبت به سایر روش‌ها بیشتر از ۵ درصد در کاهش ویژگی، بهبود داشته و در میزان صحت دسته‌بندی نسبت به بهترین نتیجه روش‌های دیگر، به‌طور میانگین در حدود ۲ درصد بهبود داشته باشد.

تاریخ دریافت:
۱۴۰۰/۰۹/۳۰

تاریخ پذیرش:
۱۴۰۱/۰۳/۱۰

کلیدواژه‌ها:
الگوریتم سینوس کسینوس، انتخاب ویژگی، بهینه‌سازی چندگانه، انتخاب ویژگی رپر

نویسنده مسئول:
saadatjou@sau.ac.ir

doi : 10.22034/ABMIR.2022.2701

۱- مقدمه

بسترهای داده‌ای که دارای ابعاد زیادی هستند علیرغم فرصت‌هایی که به وجود می‌آورند، چالش‌های محاسباتی زیادی را ایجاد می‌کنند. یکی از مشکلات داده‌های با ابعاد زیاد این است که در بیشتر مواقع تمام ویژگی‌های داده‌ها برای یافتن دانشی که در داده‌ها نهفته است مهم و حیاتی نیستند. به همین دلیل در بسیاری از زمینه‌ها کاهش ابعاد داده یکی از مباحث قابل‌توجه باقی‌مانده است و یکی از راه‌های کاهش ابعاد انتخاب ویژگی است. برای مساله انتخاب ویژگی، راه‌حل‌ها و الگوریتم‌های فراوانی ارائه‌شده است که بعضی از آن‌ها قدمت سی یا چهل‌ساله دارند. مشکل بعضی از الگوریتم‌ها درزمانی که ارائه‌شده بودند، بار محاسباتی زیاد آن‌ها بود، اگرچه امروزه با ظهور کامپیوترهای سریع و منابع ذخیره‌سازی بزرگ این مشکل، به چشم نمی‌آید ولی از طرف دیگر، مجموعه‌های داده‌های بسیار بزرگ برای مسائل جدید باعث شده است که همچنان پیدا کردن یک الگوریتم سریع برای این کار مهم باشد. مسئله انتخاب ویژگی از مسائل سخت (NP-hard) است که روش‌های کلاسیک راهکار مناسبی برای حل آن نیست زیرا هزینه محاسباتی بالایی داشته و ممکن است به بهترین زیرمجموعه ویژگی‌ها نرسند و از طرفی الگوریتم‌های تکاملی راهکار مناسبی برای حل مسئله انتخاب ویژگی هستند که با هزینه محاسبات منـاسب به بهترین مجموعه ویژگی دست پیـدا می‌کـنند.

الگوریتم‌های تکاملی همیشه به یک مجموعه ویژگی ثابت نمی‌رسند بنابراین مسئله انتخاب ویژگی یک مسئله چندگانه است و نیاز برای به دست آمدن تمام بهینه‌ها وجود دارد. درواقع انتخاب ویژگی فرآیند انتخاب زیردسته‌ای از میان دسته ویژگی‌ها است و هدف پیدا کردن یک زیرمجموعه با حداقل اندازه ممکن، برای ویژگی‌ها است، که برای هدف موردنظر اطلاعات لازم و کافی را در برداشته باشد.

از دیدگاه محاسباتی مسئله انتخاب ویژگی یک مسئله بهینه‌سازی دشوار است.

تاکنون روش‌های بهینه‌سازی بسیاری برای این مسئله پیشنهادشده‌اند که از بین آن‌ها روش‌های تکاملی مبتنی بر جمعیت توجه خاصی را به خود جلب کرده‌اند، نشان داده‌شده است که الگوریتم سینوس کسینوس به‌عنوان روش بهینه‌سازی جدید دارای دقت بالا در پیدا کردن بهینه مسائل حتی در توابع با ابعاد بالا است و از طرفی مسئله انتخاب ویژگی یک مسئله بهینه‌سازی با ابعاد بالا است که در آن هر ویژگی یک بعد مسئله است؛ و دیگر آن‌که مسئله انتخاب ویژگی یک مسئله چندگانه است زیرا در اجرای مختلف روش‌ها، همیشه یک مجموعه واحد از ویژگی‌ها به دست نمی‌آید ازاین‌رو به نظر می‌رسد الگوریتم سینوس کسینوس می‌تواند دقت پیدا کردن بهترین مجموعه ویژگی را بالابرده و همچنین بتواند با تغییراتی که در ساختار آن داده می‌شود تمام مجموعه ویژگی‌های مناسب را در اختیار قرار دهد. بنابراین در این پژوهش برای اولین بار الگوریتم سینوس کسینوس به‌واسطه دقت مناسب آن دررسیدن به بهینه مسائل، در بهینه‌سازی چندگانه به‌صورت گسسته، برای حل مسئله انتخاب ویژگی بهبود داده می‌شود.

۲- معرفی الگوریتم سینوس کسینوس

الگوریتم بهینه‌سازی سینوس کسینوس با منبع الهام از حرکت امواج سینوس و کسینوسی مدل‌سازی و توانسته نشان دهد که می‌تواند هر دو مسئله اکتشاف و استخراج را در بهینه‌سازی داشته باشد و به بهینه سرا سری مسئله دست پیدا کند. این الگوریتم دارای مکانیسم استخراج قوی است زیرا همیشه بهترین جواب نشان‌دهنده مقصد برای موج‌های جستجو است بنابراین موج‌های جستجو از بهینه اصلی مسئله منحرف نمی‌شوند. درواقع رفتار نوسانی در الگوریتم سینوس کسینوس به آن اجازه می‌دهد تا فضای جستجو را در اطراف بهینه مسئله در ابعاد بالا به‌خوبی جستجو نماید و دقت به دست آوردن بهینه خوبی داشته باشد (Mirjalili,2016). مراحل این الگوریتم در الگوریتم (۱) آمده است:

مراحل این الگوریتم در الگوریتم (۱) آمده است:

شروع: تولید موج‌های جستجو

انجام مراحل زیر تا وقتی‌که به تعداد تکرار نهایی برسد:

۱- ارزیابی هر موج با استفاده از تابع برازش مسئله

۲- به‌روزرسانی بهترین جواب پیداشده

۳- به‌روزرسانی r_1

٤- به‌روزرسانی مختصات جدید موج‌ها

$$x_i^{t+1} = \begin{cases} x_i^t + r_1 \times sin(r_2) \times |r_3 p_i^t - x_i^t| & r_4 < 0.5 \\ x_i^t + r_1 \times cos\ (r_2) \times |r_3 p_i^t - x_i^t| & r_4 \geq 0.5 \end{cases}$$

اتمام تعداد تکرار نهایی الگوریتم و برگرداندن بهترین جواب پیداشده به‌عنوان بهینه مسئله.

الگوریتم (۱): شبه کد الگوریتم بهینه‌سازی سینوس کسینوس

در این الگوریتم متغیرها شامل:

x_i^{t+1}: مختصات جدید موج جستجو

x_i^t: مختصات قبلی موج جستجو

p_i^t: مختصات بهترین جواب پیداشده که به‌عنوان مقصد در نظر گرفته می‌شود.

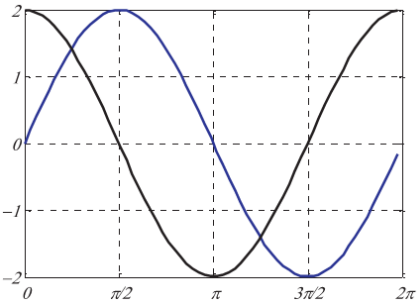
r_1: درصورتی‌که کمتر ۱ باشد به‌طرف نقطه مقصد حرکت انجام می‌شود و اگر بزرگ‌تر از ۱ باشد به‌طرف دور از نقطه مقصد حرکت انجام می‌شود.

نحوه به‌روزرسانی r_1 به‌صورت (۱) است:

$$r_1 = a - t\frac{a}{T} \quad (۱)$$

که در آن،a متغییر ثابت و t تکرار فعلی و T تکرار نهایی است.

r_2 برای مدل‌سازی حرکت نوسانی است و بین زاویه ۰ تا ۳۶۰ درجه متغیر است. تأثیر پارامتر r_2 بر حرکت امواج در الگوریتم سینوس کسینوس در شکل (۱) آمده است.



شکل (۱): تأثیر پارامتر r۲ بر حرکت امواج در الگوریتم سینوس کسینوس

r_3 یک وزن برای مقصد است و درصورتی‌که بزرگ‌تر از ۱ در نظر گرفته شود گام حرکتی بزرگ‌تری به سمت مقصد انجام می‌شود و اگر کمتر از ۱ باشد حرکت با گام کوچک‌تری انجام می‌شود.r_4 یک متغییر تصادفی بین ۰ و ۱ برای سوییچ کردن بین حرکت سینوسی و کسینوسی است.

۳- روش پیشنهادی

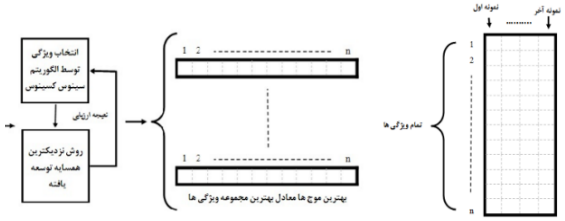
الگوریتم سینوس کسینوس (Mirjalili, 2016) نشان داده است که دارای دقت همگرایی در توابع تست با ابعاد مختلف نسبت به سایر روش‌ها است. روش نزدیک‌ترین همسایه توسعه‌یافته نیز (Harrison, 2018)برتری خود را ازنظر دقت نسبت به روش نزدیک‌ترین همسایه نشان داده است

نوآوری روش پیشنهادی در این مقاله ارائه روش انتخاب ویژگی در قالب مدل رپر با ترکیب الگوریتم سینوس کسینوس و روش نزدیک‌ترین همسایه توسعه‌یافته است) الگوریتم نزدیک‌ترین همسایه توسعه‌یافته (ENN) در مدل رپر جهت دسته‌بندی به‌کاررفته است .(در واقع تاکنون مدل رپر با استفاده از روش نزدیک‌ترین همسایه توسعه‌یافته ارائه نشده است).

همچنین نوآوری دوم ما در این مقاله ارائه مدل انتخاب ویژگی به‌صورت چندگانه است (برای اولین بار الگوریتم سینوس کسینوس به‌صورت چندگانه (Multimodal) در انتخاب ویژگی به کار گرفته‌شده است).

روش پیشنهادی دارای دو مرحله با عنوان مرحله انتخاب ویژگی و مرحله یادگیری و آزمایش است و درواقع بر اساس مدل رپر عمل می‌کند. در مرحله انتخاب ویژگی که در آن زیرمجموعه‌هایی از مجموعه ویژگی اصلی، بااستفاده از دقت طبقه‌بند سینوس کسینوس روی‌داده‌های آموزش به‌عنوان معیار، انتخاب می‌شوند. در مرحله یادگیری و آزمایش، از طبقه بند نزدیک‌ترین همسایه توسعه‌یافته با استفاده از بهترین زیرمجموعه ویژگی یافته شده و داده‌های آموزش، یاد گرفته می‌شود و با استفاده از داده‌های آزمایش محک زده می‌شود.

اگر ساختار یک مجموعه داده را به این صورت در نظر بگیریم که هر سطر نشان‌دهنده ویژگی باشد، مدل کلی روش پیشنهادی در شکل (۳) نشان داده‌شده است:



شکل (۲): شمای کلی روش پیشنهادی

که در آن برای هر موج یک بردار در نظر گرفته می‌شود که هر خانه آن مترادف با شماره ویژگی در مجموعه داده‌هاست که در شکل (۲) نمایش داده‌شده است.

در این روش، هر خانه از موج که مقدار ۱ داشته باشد بدین معنی است که آن ویژگی در مجموعه ویژگی انتخاب قرار دارد. با توجه به اینکه درروش پیشنهادی، اپراتورهای الگوریتم سینوس کسینوس بایستی به شکل باینـری باشند و با توجه به سابقه به‌کارگیـــری الگـوریتم سینـوس کسینوس در انتخـاب ویـژگی درروش پیشنهادی نیز از همان راهکار بهره برده شده است. البته درروش پیشنهادی ساختاری جهت استخراج مجموعه ویژگی‌ها در نظر گرفته‌شده است تا بدین شکل تمام مجموعه ویژگی‌های مناسب استخراج شود. برای ارزیابی هر مجموعه موج (مجموعه ویژگی) باید میزان برازندگی یا همان تابع برازندگی تخصیص یابد که برای این کار از روش نزدیک‌ترین همسایه توسعه‌یافته برای ارزیابی مجموعه ویژگی‌های انتخابی با استفاده از رابطه (۲) استفاده می‌شود (K Hussein,2021):

$$\text{fitness}_i(\text{selectedFeatures}) = \propto \left(\text{classificationAccuracy}(\text{selectedFeatures})\right) + (1-\propto)\left(\frac{N_t - N_s}{N_t}\right) \quad (۲)$$

که در آن N_t و N_s به ترتیب نماینده تعداد کل ویژگی‌ها و تعداد ویژگی‌های انتخاب‌شده و یک ضریب بین ۰ و ۱است که در شبیه‌سازی‌ها ۰.۵ در نظر گرفته‌شده است.

میزان دقت طبقه‌بندی در تابع برازندگی از رابطه (۳) به دست می‌آید:

$$CA = \frac{TN + TP}{TN + FN + TP + FP} \quad (۳)$$

که در آن:

TN: بیانگر تعداد رکوردهایی است که دسته واقعی آن‌ها منفی بوده و الگوریتم دسته‌بندی نیز دسته آن‌ها را به‌درستی منفی تشخیص داده است.

TP: بیانگر تعداد رکوردهایی است که دسته واقعی آن‌ها مثبت بوده و الگوریتم دسته‌بندی نیز دسته آن‌ها را به‌درستی مثبت تشخیص داده است.

FP: بیانگر تعداد رکوردهایی است که دسته واقعی آن‌ها منفی بوده و الگوریتم دسته‌بندی دسته آن‌ها را به‌اشتباه مثبت تشخیص داده است.

FN: بیانگر تعداد رکوردهایی است که دسته واقعی آن‌ها مثبت بوده و الگوریتم دسته‌بندی دسته آن‌ها را به‌اشتباه منفی تشخیص داده است.

فلوچارت روش پیشنهادی در شکل (۳) آمده است. همان‌طور که مشخص است جهت استخراج بهینه‌های مسئله از تکنیک سرگردانی که مبتنی بر چند جمعیت است استفاده‌شده است.

در تکنیک سرگردانی (Nekouie,2016) چندین جمعیت به دنبال بهینه‌های مسئله می‌روند و بهینه پیداشده درصورتی‌که تکراری نباشد در حافظه ذخیره می‌شود (برای تبدیل حالت مساله به حالت چندگانه، از تکنیک سرگردانی استفاده می‌شود؛ درواقع فضای مساله به چند جمعیت تقسیم‌شده و در هرکدام از این زیرجمعیت ها به دنبال بهینه بوده که این بهنیه در صورت عدم تکرار در حافظه ذخیره می‌شوند).

٤- مراحل الگوریتم روش پیشنهادی

مرحله ۱: تنظیم پارامترهای الگوریتم پیشنهادی و مسئله بهینه‌سازی شامل تعداد زیر جمعیت‌ها و تعداد موج‌ها در هر زیر جمعیت، تعداد تکرار داخلی زیر جمعیت‌ها و تعداد تکرار کلی الگوریتم.

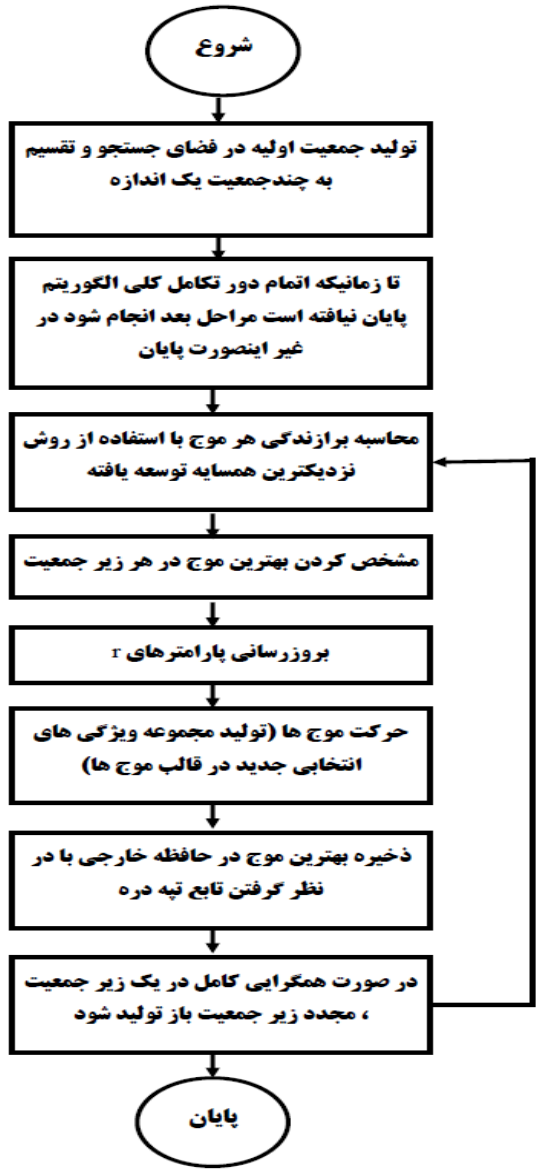
مرحله ۲: در هر زیر جمعیت تا رسیدن به اتمام تکرار داخلی مراحل زیر انجام شود.

- محاسبه برازندگی هر موج با معادله (۲) و استفاده از روش نزدیک‌ترین همسایه توسعه‌یافته.
- مشخص کردن بهترین موج (موجی با برازندگی بهتر) در هر زیر جمعیت.
- به‌روزرسانی: r_1 بروز شود.

$$r_1 = a - t\frac{a}{T} \quad (۱)$$

a متغیر ثابت، t تکرار فعلی و T تکرار نهایی.

- به‌روزرسانی: مختصات جدید موج‌ها بروز شود.
- تبدیل موج‌ها با مقادیر حقیقی به معادل باینری
- درصورتی‌که موجی از فضای جستجو خارج‌شده است، مجدد تولید شود.



شکل (۳): گراف روش سینوس کسینوس چندگانه

مرحله ۳: اضافه کردن به شمارنده تکرار داخلی و ذخیره بهترین موج در هر زیر جمعیت در حافظه خارجی (در حافظه موج‌هایی که ازنظر ساختار با موج‌های دیگر ذخیره‌شده متفاوت هستند ذخیره می‌شود) در حافظه تنها جواب‌هایی ذخیره می‌شود که تکراری نباشد.

مرحله ٤: بازتولید زیر جمعیت‌های موج‌ها.

مرحله ۵: به شمارنده اصلی اضافه می‌کنیم؛ درصورتی‌که تکرار الگوریتم تمام شود به مرحله پایان می‌رویم، در غیر این صورت به مرحله ۲ برمی‌گردیم.

۵- ارزیابی روش پیشنهادی

این الگوریتم با روش‌های بکار گرفته‌شده در (kamyab,2015). شامل روش‌هایی مبتنی بر روش PSO بانام‌های r2PSO، r3PSO، r2PSO-lhc، r3PSO-lhc و نیز روش تسهیم شایستگی پویا (DFS) که مبتنی بر روش وراثتی مقایسه می‌گردد. در این شبیه‌سازی از یک روش تکاملی که قادر به بهینه‌سازی چندگانه نیست نیز در مقایسه نتیجه‌ها استفاده‌شده است. روش یا الگوریتم CatfishBPSO به‌عنوان یک الگوریتم مکاشفه‌ای قوی موجود برای انتخاب بهترین مجموعه ویژگی در نظر گرفته‌شده است.

در تمام الگوریتم‌های جستجو تابع شایستگی که در نظر گرفته‌شده است مطابق با معادله (۲) است که ضمن حداکثر کردن دقت طبقه بند نزدیک‌ترین همسایه سعی در کم کردن تعداد ویژگی‌های موجود در زیرمجموعه ویژگی دارد.

درروش ارائه‌شده، تعداد جمعیت به تعداد زیرمجموعه از ویژگی‌هایی که قرار است به دست آورده شود مربوط نیست زیرا در هر زیر جمعیت هنگامی‌که همگرایی رخ دهد، دوباره زیر جمعیت در فضای جستجو پخش می‌شود تا مجدداً به زیرمجموعه از ویژگی‌های دیگر همگرایی شود.

درواقع ممکن است یک زیر جمعیت به مجموعه ویژگی‌های مختلفی همگرا شود و یا حتی ممکن است در این همگرایی‌ها، یک زیر جمعیت به یک مجموعه از ویژگی چند ارزش همگرا شود ولی با نظارتی که روی ذخیره داده‌ها در حافظه در الگوریتم پیشنهادی وجود دارد از ذخیره جواب‌های مشابه اجتناب می‌شود.

در طول روند جستجو برای به دست آوردن دقت طبقه‌بندی برای محاسبه شایستگی از ٦٠٪ داده‌ها به‌عنوان داده‌های آموزشی استفاده‌شده است و در پایان برای محاسبه دقت طبقه‌بندی حاصل،

از ٤٠٪ داده‌ها به‌عنوان داده آزمایش استفاده‌شده است. همچنین برای گزارش دقت دسته‌بندی برای تمام مجموعه داده‌ها، الگوریتم نزدیک‌ترین همسایه با یک همسایه استفاده‌شده است.

تنظیمات پارامترهای روش‌های مورد مقایسه به‌صورت زیر است:

- الگوریتم‌های بهبودیافته PSO مانند r2PSO، r3PSO، r2PSO-lhc، r3PSO-lhc و CatfishBPSO دارای سایز جمعیت ۷۰ با تکرار نهایی ۲۰۰ و نیز مقادیر پارامترهای دیگر ازجمله V_{min}=-6 و Vmax=6 و C_1=C_2=2 و W=1 و $\emptyset = 4.1$ مطابق (kamyab,2015) انتخاب گشته است.
- الگوریتم تسهیم شایستگی پویا (DFS) هم با جمعیت اولیه ۷۰ و تعداد تکرار ۲۰۰ و نوع گزینش چرخ رولت و تقاطع یک نقطه‌ای با احتمال ۰.۹ و احتمال جهش ۰.۰۱ بوده‌اند (kamyab,2015). همچنین تنظیمات الگوریتم پیشنهادی برای مقایسه بهتر ازنظر تعداد جمعیت اولیه برابر ۷۰ و تعداد تکرار ۲۰۰ در نظر گرفته‌شده است.

از مجموعه داده (kamyab,2015) جهت مقایسه استفاده‌شده که مشخصات آن به شرح زیر است:

- مجموعه داده شیشه یا (Glass) دارای ۲۱٤ نمونه و با تعداد ۹ ویژگی است این مجموعه داده، داده گمشده ندارد و نوع داده‌ها به‌صورت اعداد حقیقی است. این مجموعه داده دارای ۷ کلاس است.
- مجموعه داده شراب یا (Wine) دارای ۱۷۸ نمونه و با تعداد ۱۳ ویژگی است این مجموعه داده، داده گمشده ندارد و نوع داده‌ها به‌صورت اعداد حقیقی و صحیح است. این مجموعه داده دارای ۳ کلاس است.
- مجموعه داده باغ‌وحش یا (Zoo) دارای ۱۰۱ نمونه و با تعداد ۱٦ ویژگی است این مجموعه داده، داده گمشده ندارد و نوع داده‌ها به‌صورت اعداد صحیح و اسمی است. این مجموعه داده دارای ۷ کلاس است.
- مجموعه داده وسیله نقلیه یا (Vehicle) دارای ۹٤٦ نمونه و با تعداد ۱۹ ویژگی است این مجموعه داده، داده گمشده ندارد و نوع داده‌ها به‌صورت اعداد صحیح است. این مجموعه داده دارای ٤ کلاس است.
- مجموعه داده سویا یا (Soybean) دارای ۳۰۷ نمونه و با تعداد ۳۵ ویژگی است این مجموعه داده، داده گمشده دارد و نوع داده‌ها به‌صورت اسمی است. این مجموعه داده دارای ۱۹ کلاس است.
- مجموعه داده سرطان ریه یا (lung cancer) دارای ۳۲ نمونه و با تعداد ٥٦ ویژگی است این مجموعه داده، داده گمشده دارد و نوع داده‌ها به‌صورت اعداد صحیح است. این مجموعه داده دارای ۲ کلاس است.
- مجموعه داده امواج سونار یا (Sonar) دارای ۲۰۸ نمونه و با تعداد ٦۰ ویژگی است این مجموعه داده، داده گمشده ندارد و نوع داده‌ها به‌صورت اعداد حقیقی است. این مجموعه داده دارای ۲ کلاس است.
- جدول (۱) نتایج به‌دست‌آمده در بخش جستجوی رپر با طبقه بند نزدیک‌ترین همسایه را نشان می‌دهد که در آن نتایج حاصل از میانگین ۲۰ اجرای مستقل است. این نتایج شامل دقت به‌دست‌آمده از پاسخ‌های یافته شده توسط الگوریتم‌ها و داده‌های مورد آزمایش هستند. اعداد داخل پرانتز نشان‌دهنده انحراف معیار می‌باشند.

جدول (۱): نتایج حاصل از انتخاب بهترین مجموعه ویژگی با استفاده از الگوریتم‌های بهینه‌سازی چندگانه و مقایسه با الگوریتم پیشنهادی و دقت به‌دست‌آمده با استفاده از دسته‌بند نزدیک‌ترین همسایه

پایگاه داده	الگوریتم	تعداد ویژگی	دقت دسته‌بندی
Glass	بدون انتخاب ویژگی	9	0.67(0.05)
	DFS (kamyab,2015)	3.4(0.69)	0.71(0.04)
	r2PSO(kamyab,2015)	3.4(0.50)	0.71(0.04)
	r3PSO(kamyab,2015)	3.35(0.51)	0.71(0.03)
	r2PSO-lhc(kamyab,2015)	3.5(0.50)	0.78 (0.04)

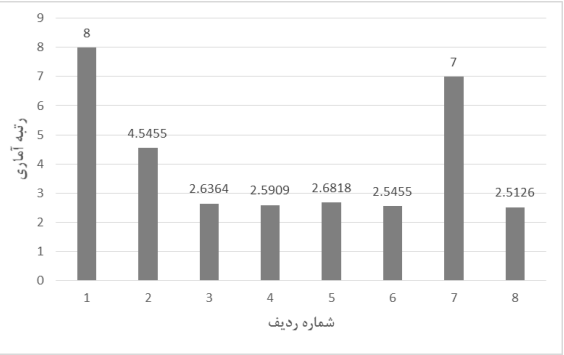
	r3PSO-lhc(kamyab,2015)	3.85(0.44)	0.74 (0.04)
	CatfishBPSO(kamyab,2015)	3.8(0.74)	0.69 (0.00)
	الگوریتم پیشنهادی	2.63(0.55)	0.98(0.02)
Wine	بدون انتخاب ویژگی	13	0.73(0.03)
	DFS (kamyab,2015)	3.45(0.43)	0.71(0.02)
	r2PSO (kamyab,2015)	2.7(0.47)	0.95(0.01)
	r3PSO(kamyab,2015)	2.55(0.54)	0.95(0.02)
	r2PSO-lhc (kamyab, 2015)	2.65(0.48)	0.95(0.02)
	r3PSO-lhc (kamyab, 2015)	2.75(0.12)	0.96(0.01)
	CatfishBPSO(kamyab, 2015)	4.12(0.63)	0.91(0.01)
	الگوریتم پیشنهادی	3.29(0.23)	0.96(0.02)
Zoo	بدون انتخاب ویژگی	16	0.93(0.06)
	DFS (kamyab,2015)	4.65(0.86)	0.96(0.03)
	r2PSO (kamyab, 2015)	3.95(0.81)	0.97(0.02)
	r3PSO (kamyab, 2015)	4.55(0.67)	0.96(0.02)
	r2PSO-lhc(kamyab,2015)	4.75(0.79)	0.97(0.03)
	r3PSO-lhc(kamyab,2015)	4.77(0.80)	0.97(0.06)
	CatfishBPSO(kamyab, 2015)	5.3(1.32)	0.97(0.01)
	الگوریتم پیشنهادی	4.05(0.26)	0.97(0.01)
Vehicle	بدون انتخاب ویژگی	19	0.69(0.01)
	DFS (kamyab,2015)	5.61(0.55)	0.73(0.01)
	r2PSO(kamyab,2015)	4.85(0.51)	0.72(0.02)
	r3PSO(kamyab,2015)	4.82(0.68)	0.73(0.08)
	r2PSO-lhc(kamyab,2015)	5.15(0.41)	0.73(0.01)
	r3PSO-lhc(kamyab,2015)	5.05(0.51)	0.73(0.04)
	CatfishBPSO(kamyab,2015)	5.45(1.52)	0.72(0.01)
	الگوریتم پیشنهادی	3.95(0.23)	0.72(0.01)
Soybean	بدون انتخاب ویژگی	35	0.98(0.03)
	DFS (kamyab,2015)	4.8(0.89)	1(0)
	r2PSO(kamyab,2015)	1.96(0.37)	1(0)
	r3PSO(kamyab,2015)	2(0)	1(0)
	r2PSO-lhc(kamyab,2015)	2(0)	1(0)
	r3PSO-lhc(kamyab,2015)	1.95(0.22)	1(0)
	CatfishBPSO (kamyab,2015)	2.25(2.66)	1(0)
	الگوریتم پیشنهادی	1.95(0.10)	1(0)
lung cancer	بدون انتخاب ویژگی	56	0.46(0.13)
	DFS (kamyab,2015)	19.4(3.01)	0.92(0.04)
	r2PSO (kamyab, 2015)	8.6(2.52)	0.98(0.03)
	r3PSO (kamyab,2015)	7.9(2.14)	0.98(0.02)
	r2PSO-lhc(kamyab,2015)	13.2(2.91)	0.94(0.05)
	r3PSO-lhc(kamyab,2015)	11.35(2.72)	0.98(0.06)

	CatfishBPSO (kamyab,2015)	25.85(4.52)	0.95(0.06)
	الگوریتم پیشنهادی	9.59(1.98)	0.95(0.02)
Sonar	بدون انتخاب ویژگی	60	0.81(0.02)
	DFS (kamyab,2015)	18.5(2.56)	0.91(0.02)
	r2PSO (kamyab,2015)	12.55(1.57)	0.96(0.02)
	r3PSO (kamyab,2015)	11.5(1.70)	0.96(0.02)
	r2PSO-lhc(kamyab,2015)	14.3(1.68)	0.95(0.01)
	r3PSO-lhc(kamyab,2015)	12.85(1.46)	0.94(0.01)
	CatfishBPSO (kamyab,2015)	16.45(3.64)	0.95(0.02)
	الگوریتم پیشنهادی	14.6(1.021)	0.96(0.01)

با توجه به نتایج به‌دست‌آمده در جدول (۱) درمی‌یابیم که خروجی الگوریتم‌های بهینه‌سازی چندگانه، مجموعه‌هایی با تعداد کمتر ویژگی هستند (کاهش ابعاد داده‌ها را در بردارند). هرچند الگوریتم CatfishBPSO توانسته در بعضی موارد به‌دقت بالاتری دست یابد لیکن تعداد ویژگی‌هایی که توسط الگوریتم CatfishBPSO حاصل می‌شود بیش از الگوریتم‌های بهینه‌ساز چندگانه است.

به‌منظور مقایسه درست‌تر، به‌طور آماری شایستگی حاصل‌شده با بهره‌گیری از نتایج موجود در جدول (۱) طبق رابطه (۲) قیاس شده‌اند تا هم توجه طبقه‌بندی و هم تعداد ویژگی حاصل‌شده در اختصاص رتبه به الگوریتم‌های موجود نقش داشته باشند. به این صورت رتبه به‌دست‌آمده به‌واسطه هر الگوریتم با بهره‌مندی از تست آماری فردمن (Pereira, 2015) با ٦ درجه آزادی مطابق با توزیع chisquare و میزان P مساوی با صفر، در شکل (۵) آورده شده است.

در شکل (٤) نمودار میله‌ای رتبه‌بندی فردمن مرتبط با شکل (۵) آمده است که در این نمودار عدد کمتر نشان‌دهنده رتبه بیشتر است.



شکل (٤): نمودار میله‌ای رتبه‌بندی فردمن مقایسه اول

رتبه‌های موجود در شکل (۵) بیانگر آن است که روش‌های بهینه‌ساز چندگانه توانسته‌اند در پیدا کردن زیرمجموعه ویژگی‌های کوچک‌تر و با توجه بیشتر، شایسته‌تر از الگوریتم CatfishBPSO عمل نمایند و نیز این الگوریتم توانسته شایسته‌ترین رتبه را از دیدگاه اندازه زیرمجموعه ویژگی‌ها و توجه و دقت دسته‌بندی داشته باشد.

در بین الگوریتم‌های بهینه‌ساز چندگانه الگوریتم‌های مبتنی بر روش PSO معرفی‌شده‌همانند الگوریتم پیشنهادی دارای نتایج بسیار مطلوبی است. ازاین‌رو قدرت اکتشاف زیاد الگوریتم‌های بهینه‌ساز چندگانه در بهبود حل مسئله انتخاب بهترین مجموعه ویژگی می‌تواند مؤثر واقع گردد. در گام بعدی دقت و توجه حاصل از طبقه‌بندی داده‌ها با بهره‌گیری از بهترین مجموعه ویژگی یافته شده توسط الگوریتم ماشین بردار پشتیبان (SVM) و درخت J48 اقدام شده است. به‌منظور این شبیه‌سازی از نرم‌افزار WEKA استفاده گشته است.

در شکل (۵) دقت و توجه حاصل از طبقه‌بندی داده‌ها با بهره‌مندی از بهترین مجموعه ویژگی یافته شده توسط الگوریتم دسته‌بندی SVM با هسته پایه شعاعی ذکر و نتایج حاصله در جدول (۲) درج گردید.

جدول (۲): رتبه آماری به‌دست‌آمده در تست فردمن با استفاده از دقت دسته‌بند SVM

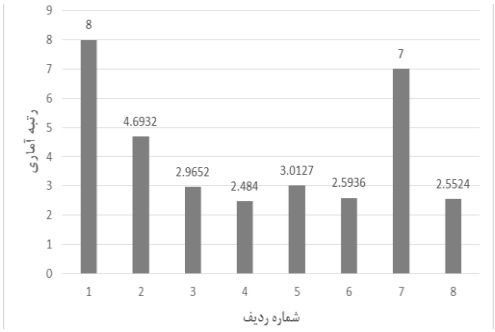
ردیف	الگوریتم	رتبه آماری
۱	بدون انتخاب ویژگی	8

۲	DFS(kamyab,2015)	4.69
۳	r2PSO(kamyab,2015)	2.96
٤	r3PSO(kamyab,2015)	2.48
٥	r2PSO-lhc(kamyab,2015)	3.01
٦	r3PSO-lhc(kamyab,2015)	2.59
۷	CatfishBPSO(kamyab,2015)	7
۸	الگوریتم پیشنهادی	2.55

جدول (۳): نتیجه دسته‌بندی بهترین پاسخ یافته شده توسط الگوریتم‌های موجود با استفاده از دسته‌بند SVM

پایگاه داده	الگوریتم	تعداد ویژگی	دقت دسته‌بندی
Glass	بدون انتخاب ویژگی	9	68.66
	DFS (kamyab,2015)	3	72.09
	r2PSO(kamyab,2015)	3	71.01
	r3PSO(kamyab,2015)	3	71.21
	r2PSO-lhc(kamyab,2015)	4	79.56
	r3PSO-lhc(kamyab,2015)	4	75.15
	CatfishBPSO (kamyab,2015)	4	70.44
	الگوریتم پیشنهادی	3	98.85
Wine	بدون انتخاب ویژگی	13	74.22
	DFS (kamyab,2015)	3	72.14
	r2PSO(kamyab,2015)	3	95.50
	r3PSO(kamyab,2015)	3	94.42
	r2PSO-lhc(kamyab,2015)	3	95.52
	r3PSO-lhc(kamyab,2015)	3	97.52
	CatfishBPSO (kamyab,2015)	4	92.39
	الگوریتم پیشنهادی	3	98.11
Zoo	بدون انتخاب ویژگی	16	93.51
	DFS (kamyab,2015)	5	93.25
	r2PSO(kamyab,2015)	4	97.11
	r3PSO(kamyab,2015)	5	96.38
	r2PSO-lhc(kamyab,2015)	5	97.20
	r3PSO-lhc(kamyab,2015)	5	97.61
	CatfishBPSO (kamyab,2015)	5	97.11
	الگوریتم پیشنهادی	4	97.91
Vehicle	بدون انتخاب ویژگی	19	69.40
	DFS (kamyab,2015)	6	74.31
	r2PSO(kamyab,2015)	5	72.36
	r3PSO(kamyab,2015)	5	73.49
	r2PSO-lhc(kamyab,2015)	5	73.62
	r3PSO-lhc(kamyab,2015)	5	73.32

	72.04	5	CatfishBPSO (kamyab,2015)
	73.9	4	الگوریتم پیشنهادی
Soybean	98.82	35	بدون انتخاب ویژگی
	1	5	DFS (kamyab,2015)
	1	2	r2PSO(kamyab,2015)
	1	2	r3PSO(kamyab,2015)
	1	2	r2PSO-lhc(kamyab,2015)
	1	2	r3PSO-lhc(kamyab,2015)
	1	2	CatfishBPSO (kamyab,2015)
	1	2	الگوریتم پیشنهادی
lung cancer	47.61	56	بدون انتخاب ویژگی
	93.51	20	DFS (kamyab,2015)
	97.38	9	r2PSO (kamyab,2015)
	97.19	8	r3PSO (kamyab,2015)
	94.48	13	r2PSO-lhc(kamyab,2015)
	97.61	11	r3PSO-lhc(kamyab,2015)
	95.06	26	CatfishBPSO (kamyab,2015)
Sonar	97.17	10	الگوریتم پیشنهادی
	82.29	60	بدون انتخاب ویژگی
	91.22	19	DFS (kamyab,2015)
	96.01	13	r2PSO(kamyab,2015)
	97.46	11	r3PSO(kamyab,2015)
	95.10	14	r2PSO-lhc(kamyab,2015)
	94.87	13	r3PSO-lhc(kamyab,2015)
	96.43	16	CatfishBPSO (kamyab,2015)



شکل (۵): نمودار میله‌ای رتبه‌بندی فردمن مقایسه دوم

در شکل (۵) نمودار میله‌ای رتبه‌بندی فردمن مرتبط با جدول (۲) آمده است که در آن نمودار عدد کمتر نشان‌دهنده رتبه بیشتر است. در جدول (۳) توجه حاصل از طبقه‌بندی داده‌ها با بهره‌گیری از شایسته‌ترین مجموعه ویژگی یافته شده به‌واسطه الگوریتم دسته‌بندی درخت J48 آمده است. در شبیه‌سازی J48 میزان فاکتور اطمینان (Confidence Factor) به‌منظور حرص درخت (برای کاهش ویژگی‌ها) ۰.۲۵ بوده است.

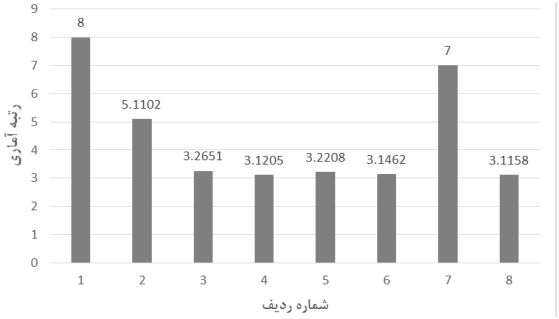
جدول (٤): رتبه آماری به‌دست‌آمده در تست فردمن

با استفاده از دقت دسته‌بندی J٤٨

ردیف	الگوریتم	رتبه آماری
۱	بدون انتخاب ویژگی	8
۲	DFS(kamyab,2015)	5.11
۳	r2PSO(kamyab,2015)	3.26
٤	r3PSO(kamyab,2015)	3.12
۵	r2PSO-lhc(kamyab,2015)	3.22

٦	r3PSO-lhc(kamyab,2015)	3.14
۷	CatfishBPSO(kamyab,2015)	7
۸	الگوریتم پیشنهادی	3.11

در شکل (٦) نمودار میله‌ای رتبه‌بندی فردمن مرتبط با جدول (٤) آمده است که در این نمودار عدد کمتر نشان‌دهنده رتبه بیشتر است. با توجه به نتایج به‌دست‌آمده با استفاده از سایر طبقه بندها مشاهده می‌کنیم که در مورد آن‌ها هم پاسخ‌های یافته شده به‌واسطه الگوریتم‌های بهینه‌ساز چندگانه با تعداد قابل‌توجه کمتر ویژگی، توانسته در بیشتر موارد حتی به‌دقت بیشتری نسبت به الگوریتم CatfishBPSO دست نماید.



شکل (٦): نمودار میله‌ای رتبه‌بندی فردمن مقایسه سوم

جدول (٥): نتیجه بهترین پاسخ یافته شده توسط الگوریتم‌های موجود

پایگاه داده	الگوریتم	تعداد ویژگی	دقت دسته‌بندی
Glass	بدون انتخاب ویژگی	9	67.65
	DFS (kamyab,2015)	3	71.09
	r2PSO(kamyab,2015)	3	70.00
	r3PSO(kamyab,2015)	3	70.20
	r2PSO-lhc(kamyab,2015)	4	78.36
	r3PSO-lhc(kamyab,2015)	4	73.03
	CatfishBPSO (kamyab,2015)	4	70.33
	الگوریتم پیشنهادی	3	96.81
Wine	بدون انتخاب ویژگی	13	73.21
	DFS (kamyab,2015)	3	72.03
	r2PSO(kamyab,2015)	3	93.39
	r3PSO(kamyab,2015)	3	93.32
	r2PSO-lhc(kamyab,2015)	3	93.30
	r3PSO-lhc(kamyab,2015)	3	96.39
	CatfishBPSO (kamyab,2015)	4	92.25
	الگوریتم پیشنهادی	3	97.01
Zoo	بدون انتخاب ویژگی	16	91.11
	DFS (kamyab,2015)	5	92.23
	r2PSO(kamyab,2015)	4	96.05
	r3PSO(kamyab,2015)	5	95.01
	r2PSO-lhc(kamyab,2015)	5	95.20
	r3PSO-lhc(kamyab,2015)	5	96.60
	CatfishBPSO (kamyab,2015)	5	96.05
	الگوریتم پیشنهادی	4	96.90
Vehicle	بدون انتخاب ویژگی	19	68.30
	DFS (kamyab,2015)	6	71.20
	r2PSO(kamyab,2015)	5	72.20
	r3PSO(kamyab,2015)	5	71.39

	r2PSO-lhc (kamyab,2015)	5	72.60
	r3PSO-lhc(kamyab,2015)	5	71.20
	CatfishBPSO (kamyab,2015)	5	72.03
	الگوریتم پیشنهادی	4	72.93
Soybean	بدون انتخاب ویژگی	35	97.80
	DFS (kamyab,2015)	5	1
	r2PSO(kamyab,2015)	2	1
	r3PSO(kamyab,2015)	2	1
	r2PSO-lhc(kamyab,2015)	2	1
	r3PSO-lhc(kamyab,2015)	2	1
	CatfishBPSO (kamyab,2015)	2	1
	الگوریتم پیشنهادی	2	1
lung cancer	بدون انتخاب ویژگی	56	36.60
	DFS (kamyab,2015)	20	87.30
	r2PSO(kamyab,2015)	9	86.28
	r3PSO(kamyab,2015)	8	91.09
	r2PSO-lhc(kamyab,2015)	13	89.38
	r3PSO-lhc(kamyab,2015)	11	93.60
	CatfishBPSO (kamyab,2015)	26	89.06
	الگوریتم پیشنهادی	10	93.07
Sonar	بدون انتخاب ویژگی	60	80.20
	DFS (kamyab,2015)	19	89.20
	r2PSO(kamyab,2015)	13	93.04
	r3PSO(kamyab,2015)	11	94.35
	r2PSO-lhc(kamyab,2015)	14	91.30
	r3PSO-lhc(kamyab,2015)	13	93.80
	CatfishBPSO (kamyab,2015)	16	95.30
	الگوریتم پیشنهادی	14	97.00

۶- جمع‌بندی و نتیجه‌گیری

انتخاب بهترین مجموعه ویژگی در یادگیری ماشین به‌منظور فرآیند انتخاب زیرمجموعه‌ای از ویژگی‌ها (ابعاد) به جهت پردازش داده استفاده می‌گردد و از آن به‌عنوان بخشی از پیش‌پردازش داده‌ها می‌توان یادکرد که به‌طور بالقوه به دنبال ساخت ویژگی یا چنانچه درروش یادگیری تعبیه گردد می‌تواند با فاز یادگیری هم همراه گردد. یک فرض در انتخاب بهترین مجموعه ویژگی آن است که می‌تواند یک فضای ویژگی اصلی برای نمایش داده‌ها بکار گرفته شود و هدف ما این است که از طریق گزینش زیرمجموعه‌ای از ویژگی‌های اصلی ابعاد آن را کاهش دهیم. پس‌ازآن فضای ویژگی اصلی داده‌ها به یک محیط ویژگی تازه نگاشت می‌شود.

در این مقاله، مدل تازه‌ای مبنی بر روش سینوس کسینوس ارائه شد که قادر باشد تمام زیرمجموعه‌های مناسب و شایسته از ویژگی را به دست آورد. به‌طور خلاصه اولین نوآوری در این مقاله ارائه الگوریتم انتخاب بهترین مجموعه ویژگی در قالب مدل رپر با ترکیب روش سینوس کسینوس و الگوریتم نزدیک‌ترین همسایه توسعه‌یافته است و دومین نوآوری، ارائه مدل انتخاب بهترین مجموعه ویژگی به‌صورت چندگانه است.

با توجه به نتایج آزمایش‌های انجام‌شده مشخص گردید که روش‌های بهینه‌سازی چندگانه توانسته‌اند مجموعه ویژگی‌هایی با تعداد کمتر ویژگی به‌منظور پایگاه‌های داده مختلف پیدا نمایند. اگرچه الگوریتم غیر چندگانه هم توانسته است در برخی موارد به‌دقت بیشتری دست یابد لیکن تعداد ویژگی‌هایی که انتخاب‌شده است، از الگوریتم‌های بهینه‌ساز چندگانه بیشتر است. می‌بینیم که برای مقایسه درست‌تر، با دقت به اینکه هم توجه و دقت طبقه‌بندی

و هم تعداد ویژگی حاصل‌شده در اختصاص رتبه به روش‌های موجود نقش داشته باشند، رتبه‌بندی تست آماری فِردمن با ٦ درجه آزادی اعمال شد. به‌این‌ترتیب رتبه حاصل‌شده توسط هر الگوریتم با بهره‌مندی از این تست نشان می‌دهد که الگوریتم‌های بهینه‌ساز چندگانه توانسته‌اند در پیدا کردن زیرمجموعه ویژگی‌های کوچک‌تر و با دقت بیشتر، با موفقیت بیشتری از الگوریتم‌های غیر چندگانه عمل نمایند و نیز دراین‌بین الگوریتم پیشنهادی توانسته شایسته‌ترین رتبه را ازنظر اندازه زیرمجموعه ویژگی‌ها و دقت و توجه دسته‌بندی داشته باشد.

در بین الگوریتم‌های بهینه‌ساز چندگانه الگوریتم‌های مبتنی بر روش PSO معرفی‌شده نیز همانند الگوریتم پیشنهادی دارای نتایج بسیار خوبی بودند. بنابراین، این موضوع دریافت شد که قدرت اکتشاف بالای الگوریتم‌های بهینه‌ساز چندگانه می‌تواند در بهبود حل مسئله انتخاب بهترین مجموعه ویژگی مؤثر واقع شود.

با عنایت به نتایج به‌دست‌آمده با بهره‌گیری از سایرطبقه‌بندها مشاهده می‌کنیم که در مورد آن‌ها هم جواب‌های کشف‌شده توسط الگوریتم‌های بهینه‌ساز چندگانه در ضمن داشتن تعداد قابل‌توجه کمتر ویژگی، توانسته در اکثر موارد حتی به‌دقت بیشتری نسبت به الگوریتم CatfishBPSO که یک الگوریتم غیر چندگانه است، دست پیدا نمایند. ازاین‌رو می‌توان بیان کرد الگوریتم‌های بهینه‌سازی چندگانه می‌توانند به‌عنوان ابزار مناسبی برای حل مسئله انتخاب بهترین مجموعه ویژگی به کار گرفته شوند و نتیجه الگوریتم پیشنهادی نیز بیانگر آن است که در بهینه‌سازی چندگانه و در کاربرد انتخاب بهترین مجموعه ویژگی، توانسته به‌خوبی استفاده گردد. لازم به ذکر است که نتایج الگوریتم پیشنهادی نسبت به الگوریتم‌های دیگر چندگانه به علت وجود الگوریتم نزدیک‌ترین همسایه توسعه‌یافته در خود، دارای نتایج بهتر بوده است.

در این مقاله به‌منظور نشان دادن کارایی و اثربخشی الگوریتم پیشنهادی عملکرد آن روی هفت مجموعه داده متفاوت از بعد تعداد نمونه و تعداد ویژگی‌ها، آزمایش شد تا بدین‌صورت مقایسه مناسبی در کنار نتایج بقیه الگوریتم‌ها انجام گردد. در کل می‌توان نتیجه گرفت که رویکرد الگوریتم‌های چندگانه نسبت به الگوریتم‌های تک‌گانه در مسئله انتخاب بهترین مجموعه ویژگی در حالت کلی دارای عملکرد بسیار شایسته‌تری بوده‌اند.

References

[1] Abualigah, L. M., & Dulaimi, A. J. (2021). A novel feature selection method for data mining tasks using hybrid Sine Cosine Algorithm and Genetic Algorithm. Cluster Computing, 24(3), 2161–2176.

[2] Ahmed, R., Nazir, A., Mahadzir, S., Shorfuzzaman, M., & Islam, J. (2021). Niching grey wolf optimizer for multimodal optimization problems. Applied Sciences, 11(11), 4795.

[3] Du, W., Ren, Z., Chen, A., & Liu, H. (2021). A Knowledge Transfer-Based Evolutionary Algorithm for Multimodal Optimization. In 2021 IEEE Congress on Evolutionary Computation (CEC) (pp. 1953–1960).

[4] Hancer, E., Xue, B., & Zhang, M. (2020). A survey on feature selection approaches for clustering. Artificial Intelligence Review, 53(6), 4519–4545.

[5] Hans, R., & Kaur, H. (2020). Hybrid binary Sine Cosine Algorithm and Ant Lion Optimization (SCALO) approaches for feature selection problem. International Journal of Computational Materials Science and Engineering, 9(1), 1950021.

[6] Harrison, O. (2018). Machine learning basics with the k-nearest neighbors algorithm. *Towards Data Science*, *11*.

[7] Hussain, K., Neggaz, N., Zhu, W., & Houssein, E. H. (2021). An efficient hybrid sine-cosine Harris hawks optimization for low and high-dimensional feature selection. Expert Systems With Applications, 176, 114778.

[8] Kamyab, S., & Eftekhari, M. (2016). Feature selection using multimodal optimization techniques. Neurocomputing, 171, 586–597.

[9] Kumar, L., & Bharti, K. K. (2021). A novel hybrid BPSO–SCA approach for feature selection. Natural Computing, 20(1), 39–61.

[10] Lu, H., Sun, S., Cheng, S., & Shi, Y. (2021). An adaptive niching method based on multi-strategy fusion for multimodal optimization. Memetic Computing, 13(3), 341–357.

[11] Mirjalili, S. (2016). SCA: A Sine Cosine Algorithm for solving optimization problems. Knowledge Based Systems, 96(96), 120–133.

[12] Neggaz, N., Ewees, A. A., Elaziz, M. E. A., & Mafarja, M. M. (2020). Boosting salp swarm algorithm by sine cosine algorithm and disrupt operator for feature selection. Expert Systems With Applications, 145, 113103.

[13] Nekouie, N., & Yaghoobi, M. (2016). A new method in multimodal optimization based on firefly algorithm. Artificial Intelligence Review, 46(2), 267–287.
[14] Nguyen, B. H., Xue, B., & Zhang, M. (2020). A survey on swarm intelligence approaches to feature selection in data mining. Swarm and Evolutionary Computation, 54, 100663.
[15] Pandit, A. A., Pimpale, B., & Dubey, S. (2020). A Comprehensive Review on Unsupervised Feature Selection Algorithms, 255–266.
[16] Pereira, D. G., Afonso, A., & Medeiros, F. M. (2015). Overview of Friedman's test and post-hoc analysis. *Communications in Statistics-Simulation and Computation*, *44*(10), 2636-2653.
[17] Ros, F., & Guillaume, S. (2020). From Supervised Instance and Feature Selection Algorithms to Dual Selection: A Review, 83–128.
[18] Sekhar, P. R., & Sujatha, B. (2020). A Literature Review on Feature Selection using Evolutionary Algorithms. In 2020 7th International Conference on Smart Structures and Systems (ICSSS) (pp. 1–8).
[19] Sheng, W., Wang, X., Wang, Z., Li, Q., & Chen, Y. (2021). Adaptive memetic differential evolution with niching competition and supporting archive strategies for multimodal optimization. Information Sciences, 573, 316–331.
[20] Tang, B., & He, H. (2015). ENN: Extended Nearest Neighbor Method for Pattern Recognition [Research Frontier]. IEEE Computational Intelligence Magazine, 10(3), 52–60.
[21] Wan, Y., Ma, A., Zhong, Y., Hu, X., & Zhang, L. (2020). Multiobjective Hyperspectral Feature Selection Based on Discrete Sine Cosine Algorithm. IEEE Transactions on Geoscience and Remote Sensing, 58(5), 3601–3618.

Journal of Applied And Basic Machine Intelligence Research (ABMIR)

Applied and Basic Machine Intelligence Research

Vol. 1, No. 1, Spring & Summer 2023, 46-59

Provide version of multimodal Sine Cosine Algorithm in solving feature selection problem

Fatemeh Saadatjoo [1*], Soheil Eghbali [1], Alireza Poursoleyman[1]

[1] Computer Engineering Department, Science and Arts University, Yazd, Iran.

Article Information

Original Research Paper

Received:
2022 December 21

Accepted:
2022 May 31

Keywords:
Sin Cosine Algorithm, Feature Selection, Multimodal Optimization, Wrapper Method

Corresponding Author*:
saadatjou@sau.ac.ir

Abstract

Because not all the features of the data are important and vital to find the knowledge that is hidden in them; Reducing their characteristic dimensions is one of the important topics. Therefore, in this article, a new method using the sine-cosine algorithm with a multiple optimization approach is presented in the field of feature selection. The proposed method is presented in the wrapper feature selection model (a set of evaluation function algorithms based on the classifier error rate) and has two stages, which include the feature selection stage using the multiple sine-cosine algorithm and the stage of classifying possible answers in the sine-cosine algorithm with the method the nearest neighbor is developed. The proposed method has been tested on the standard UCI dataset in datasets with different dimensions.The comparison of the proposed method with multiple and single optimization methods shows that in general, the proposed method has been able to reduce the number of features by more than 5% compared to other methods, and has improved the accuracy of the classification compared to the best results of other methods, to have an average improvement of about 2 percent.

: 10.22034/ABMIR.2022.2701