ارائه‌ی یک روش انتخاب ویژگی جدید مبتنی بر بهینه‌سازی ازدحام ذرات با استفاده از به‌روزرسانی فازی

سمیرا حیدری مقدم بجستانی[۱]، کارشناس ارشد؛ سعید شعرباف تبریزی[۲]، استادیار؛ عادل قاضی خانی[۳]، استادیار

۱- دانشکده مهندسی - گروه مهندسی کامپیوتر - دانشگاه بین‌المللی امام رضا (ع) - مشهد - ایران - s.heidarimoghaddam@imamreza.ac.ir
۲- دانشکده مهندسی - گروه مهندسی برق - دانشگاه بین‌المللی امام رضا (ع) - مشهد - ایران - shaerbaf@imamreza.ac.ir
۳- دانشکده مهندسی - گروه مهندسی کامپیوتر - دانشگاه بین‌المللی امام رضا (ع) - مشهد - ایران - aghazi@imamreza.ac.ir

چکیده: انتخاب ویژگی یکی از مسائل مهم در رده‌بندی است که نقش مهمی در افزایش کارایی دارد و روش‌های متفاوتی برای حل آن وجود دارد. بهینه‌سازی ازدحام ذرات یکی از الگوریتم‌های مبتنی بر هوش جمعی است که در زمینه‌های متفاوتی از جمله انتخاب ویژگی استفاده شده و کارایی خوبی از خود نشان داده است. پژوهش‌های بسیاری از بهینه‌سازی ازدحام ذرات برای انتخاب ویژگی استفاده نموده‌اند. در یکی از پژوهش‌های انجام شده در این زمینه، نویسندگان چندین راهبرد مختلف برای مقداردهی اولیه‌ی ذرات و چندین روش برای به‌روزرسانی بهترین تجربه‌ی شخصی و بهترین تجربه‌ی گروه در بهینه‌سازی ازدحام ذرات برای انتخاب ویژگی ارائه داده‌اند و به نتایج خوبی دست یافته‌اند. ما در این مقاله بر اساس پژوهش ذکر شده و به‌روزرسانی فازی پیشنهادی خود برای یکی از دو مورد بهترین تجربه‌ی شخصی یا بهترین تجربه‌ی گروه، روشی برای انتخاب ویژگی ارائه داده‌ایم. k نزدیک‌ترین همسایه به‌عنوان رده‌بند استفاده شده است. آزمایش‌ها بر روی چندین مجموعه داده انجام گرفته است. با توجه به شبیه‌سازی‌های انجام‌شده، روش پیشنهادی نتایج مطلوبی از لحاظ دقت و تعداد ویژگی در مقایسه با مقاله‌ی مرجع به‌دست آورده است.

واژه‌های کلیدی: انتخاب ویژگی، بهینه‌سازی ازدحام ذرات، فازی.

A New Feature Selection Method Based on Fuzzy Updated Particle Swarm Optimization

S. Heidari Moghaddam Bajestani[1], MSc; S. Shaerbaf Tabrizi[2], Assistant Professor; A. Ghazikhani[3], Assistant Professor

1- Faculty of Engineering, Department of Computer Engineering, Imam Reza International University, Mashhad, Iran, Email: s.heidarimoghaddam@imamreza.ac.ir
2- Faculty of Engineering, Department of Electrical Engineering, Imam Reza International University, Mashhad, Iran, Email: shaerbaf@imamreza.ac.ir
3- Faculty of Engineering, Department of Computer Engineering, Imam Reza International University, Mashhad, Iran, Email: aghazi@imamreza.ac.ir

Abstract: Feature selection is one of the important problems in classification that has an important role in increasing efficiency and there are different methods to solve it. Particle swarm optimization is one of the algorithms based on swarm intelligence that has been used in different contexts including feature selection and has shown good performance. Many studies have used particle swarm optimization for feature selection. In a research accomplished in the field, the authors have presented several different strategies for initialization of particles and several methods to update personal best and global best in particle swarm optimization for feature selection and have achieved good results. In this article we have presented a method for feature selection based on the mentioned research and our proposed fuzzy updating for one of the personal best or global best. *k*-nearest neighbor is used as the classifier. Experiments is performed on several datasets. According to the done simulations, the proposed method obtains good results in terms of accuracy and the number of feature in comparison with reference article.

Keywords: Feature selection, Particle swarm optimization, Fuzzy.

تاریخ ارسال مقاله: ۱۳۹۵/۱۰/۱۰
تاریخ اصلاح مقاله: ۱۳۹۵/۱۲/۱۸، ۱۳۹۶/۱۲/۲۱و ۱۳۹۶/۰۵/۲۵
تاریخ پذیرش مقاله: ۱۳۹۷/۰۲/۱۲
نام نویسنده مسئول: عادل قاضی خانی
نشانی نویسنده مسئول: ایران - مشهد - خیابان دانشگاه - خیابان اسرار - دانشگاه بین‌المللی امام رضا علیه‌السلام – دانشکده مهندسی – گروه مهندسی کامپیوتر

۱- مقدمه

مسائل رده‌بندی اغلب شامل تعداد زیادی ویژگی هستند ولی همه‌ی آن‌ها برای رده‌بندی مفید نمی‌باشند، ویژگی‌های زائد[۱] و نامربوط[۲] می‌توانند باعث کاهش دقت رده‌بندی شوند [۱]. یک مسئله‌ی مهم در یادگیری ماشین این است که چطور زیرمجموعه‌ی خوبی از ویژگی‌ها انتخاب شود [۲]. روش‌های انتخاب ویژگی زیرمجموعه‌ای از ویژگی‌هایی را انتخاب می‌کنند که با مفهوم هدف مرتبط هستند [۳]. روش‌های انتخاب ویژگی باعث کاهش زمان محاسبه، بهبود عملکرد پیش‌بینی و درک بهتر داده‌ها در یادگیری ماشین یا کاربردهای تشخیص الگو می‌شوند [۴]. روش‌های حل مسئله‌ی انتخاب ویژگی به دنبال یک راه‌حل (یک زیرمجموعه از کل ویژگی‌ها را یک راه‌حل می‌گویند) بهینه هستند؛ این در حالی است که فضای جستجوی این مسئله معمولاً بزرگ است [۵]. بنابراین برای یافتن زیرمجموعه‌ی بهینه نیاز به یک روش جستجوی سراسری کارآمد و مؤثر وجود دارد [۱]. بهینه سازی ازدحام ذرات (PSO[۳]) [۶، ۷] قادر است با جستجوی مؤثر در فضاهای بزرگ راه‌حلی را پیدا کند که بهینه یا نزدیک به بهینه است، همچنین دارای هزینه‌ی محاسباتی کم و سرعت همگرایی بالایی است [۱].

پژوهش‌های بسیاری در انتخاب ویژگی با استفاده از الگوریتم بهینه‌سازی ازدحام ذرات انجام گرفته است که در ادامه به برخی از آن‌ها اشاره شده است.

Wang و همکارانش در [۸] یک روش انتخاب ویژگی بر اساس تئوری مجموعه‌ی ناهموار[۴] و بهینه‌سازی ازدحام ذرات ارائه دادند که در آن ذرات به‌صورت دودویی نمایش داده می‌شوند. روی سرعت ذرات و همچنین کاهش وزن اینرسی کار شده است. Huang و Dun در [۹] یک الگوریتم پوشش[۵] برای انتخاب ویژگی و بهینه‌سازی پارامترها در ماشین بردار پشتیبان (SVM[۶]) با بهینه سازی ازدحام ذرات پیوسته و بهینه‌سازی ازدحام ذرات گسسته ارائه دادند. این الگوریتم، بهینه‌سازی ازدحام ذرات گسسته را با بهینه سازی ازدحام ذرات پیوسته، ترکیب می‌کند تا هم‌زمان زیرمجموعه‌ی ویژگی‌ها و پارامترهای ماشین بردار پشتیبان را بهینه کند. در کدگذاری هر ذره، بخش مربوط به ویژگی‌ها، توسط بهینه‌سازی ازدحام ذرات گسسته و بخش مربوط به پارامترها در ماشین بردار پشتیبان، توسط بهینه سازی ازدحام ذرات پیوسته انجام می‌شود. Yang و همکارانش در [۵] یک راهبرد تنظیم مجدد بهترین تجربه‌ی گروه[۷] (*gbest*) در بهینه سازی ازدحام ذرات دودویی ارائه دادند که از همگرایی ذرات در کمینه‌ی محلی جلوگیری شود و در انتخاب ویژگی به‌کار بردند. Chuang و همکارانش در [۱۰] یک راهبرد برای بهترین تجربه‌ی گروه در بهینه‌سازی ازدحام ذرات دودویی برای انتخاب ویژگی ارائه دادند که در آن اگر بهترین تجربه‌ی گروه، بعد از چند تکرار (۳ تکرار) مقدار یکسانی داشته باشد، با صفر تنظیم می‌شود. آزمایش‌ها بر روی چندین مجموعه داده‌ی بیان ژن سرطان در انسان نشان داد که این الگوریتم در اکثر موارد دقت رده‌بندی بهتری نسبت به روش‌های مقایسه شده، داشته است. Esseghir و همکارانش در [۱۱] از ترکیب روش‌های فیلتر[۸] و پوشش به همراه بهینه سازی ازدحام ذرات پیوسته برای انتخاب ویژگی استفاده کردند. مقداردهی اولیه‌ی ذرات توسط امتیاز فیلتر ویژگی‌ها انجام می‌شود. در این کار با چندین معیار فیلتر الگوریتم‌هایی ایجاد و با بهینه‌سازی ازدحام ذرات دودویی مقایسه شدند. Chuang و همکارانش در [۱۲] یک الگوریتم بهینه سازی جدید برای انتخاب ویژگی به نام CatfishBPSO ارائه دادند که برای بهبود عملکرد بهینه سازی ازدحام ذرات دودویی به‌کار می‌رود. در catfish اگر بهترین تجربه‌ی گروه برای سه تکرار تغییری نداشته باشد، ذرات جدیدی تولید می‌شوند که با ذرات دارای بدترین برازندگی[۹] جایگزین می‌شوند. راه‌حل‌های بهتر می‌توانند باعث هدایت ذرات به مناطق بهتری در فضای جستجو شوند. Unler و همکارانش در [۱۳] یک الگوریتم انتخاب ویژگی ترکیبی فیلتر و پوشش ارائه دادند. مدل فیلتر بر اساس اطلاعات متقابل[۱۰] است و مدل پوشش یک الگوریتم بهینه‌سازی ازدحام ذرات گسسته‌ی اصلاح شده است. از رده‌بند ماشین بردار پشتیبان در این کار استفاده شده است. Vieira و همکارانش در [۱۴] یک روش انتخاب ویژگی پوشش مبتنی بر بهینه‌سازی ازدحام ذرات دودویی و ماشین بردار پشتیبان برای پیش‌بینی مرگ‌ومیر در بیماران سپتیک[۱۱] به‌کار بردند که در آن هم‌زمان پارامترهای کرنل ماشین بردار پشتیبان تنظیم می شود. در این روش از همگرایی زودهنگام بهینه سازی ازدحام ذرات دودویی جلوگیری می‌شود. روشی برای تغییرپذیری[۱۲] جمع به‌کار برده شده است. Ghamisi و Benediktsson در [۱۵] یک روش انتخاب ویژگی جدید با ترکیب الگوریتم ژنتیک و بهینه‌سازی ازدحام ذرات ارائه دادند. از رده‌بند ماشین بردار پشتیبان در تابع برازندگی استفاده شده است. این روش بر روی یک مجموعه داده‌ی فراطیفی اجرا گردیده و در تشخیص جاده نیز آزمایش شده است و به نتایج مطلوبی دست یافته است. Gunasundari و Janakiraman در [۱۶] از ترکیب بهینه‌سازی ازدحام ذرات و انتخاب رو به‌جلو متوالی (SFS[۱۳]) و انتخاب رو به‌عقب متوالی (SBS[۱۴]) برای انتخاب ویژگی استفاده نمودند. در این پژوهش از بهینه سازی ازدحام ذرات دودویی استفاده شده است. بعد از تعدادی تکرار بهینه‌سازی ازدحام ذرات، ۳۰ درصد از بدترین ذرات با بهترین زیرمجموعه‌ی ویژگی الگوریتم انتخاب رو به‌جلو متوالی و انتخاب رو به عقب متوالی جایگزین می‌شود. دو ترکیب PSO-SFS و PSO-SFS-SBS با دو مورد تنظیم مجدد سرعت همه‌ی ذرات و تنظیم مجدد سرعت بدترین ذرات پیشنهاد شده است. این روش ترکیبی بر روی داده‌های سرطان کبد به‌کار برده شده است. Xue و همکارانش در [۱۷] از بهینه‌سازی ازدحام ذرات پیوسته برای انتخاب ویژگی استفاده کردند و سه راهبرد مقداردهی اولیه‌ی جدید و سه رویکرد به‌روزرسانی جدید برای بهترین تجربه‌ی شخصی[۱۵] و بهترین تجربه‌ی گروه با هدف افزایش کارایی رده‌بندی، کاهش تعداد ویژگی‌ها و کاهش زمان محاسباتی ارائه دادند و با رویکرد به‌روزرسانی و مقداردهی اولیه‌ی مرسوم مقایسه کردند. ترکیب امیدبخش‌ترین راهبرد مقداردهی اولیه و رویکرد به‌روزرسانی را PSO(4-2) نامیده‌اند و مقایسه‌هایی با چند روش دیگر

انجام داده‌اند. آزمایش‌ها بر روی ۲۰ مجموعه داده‌ی معیار انجام گرفته و به نتایج مطلوبی دست یافته است.

هدف از این پژوهش، بهبود عملکرد الگوریتم انتخاب ویژگی مبتنی بر بهینه‌سازی ازدحام ذرات ارائه‌شده توسط Xue و همکارانش در [۱۷] با فازی است که بتوان دقت رده‌بندی را افزایش و تعداد ویژگی‌ها را کاهش داد. برای این منظور از فازی در یکی از دو مورد به‌روزرسانی بهترین تجربه‌ی شخصی یا بهترین تجربه‌ی گروه استفاده شده و حالات مختلفی ارائه شده است. برای مقایسه‌ی نتایج، تعدادی از مجموعه داده‌های استفاده‌شده در [۱۷] از مجموعه داده‌ی UCI[۱۶] [۱۸] مورد استفاده قرار گرفته است. همچنین آزمایش‌هایی بر روی چندین مجموعه داده‌ی بیان ژن گرفته‌شده از [۱۹] انجام شده است.

بخش دوم مربوط به مفاهیم مرتبط است. در بخش سوم روش پیشنهادی ارائه شده است. بخش چهارم شامل نتایج آزمایش‌ها است و در بخش پنجم نتیجه‌گیری ارائه شده است.

۲- مفاهیم مرتبط

۲-۱- بهینه‌سازی ازدحام ذرات

الگوریتم بهینه‌سازی ازدحام ذرات توسط Kennedy و Eberhart در سال ۱۹۹۵ ارائه شد [۶، ۷]. بهینه‌سازی ازدحام ذرات رفتارهای اجتماعی مثل گروه ماهی‌ها و پرندگان را شبیه‌سازی می‌کند. در آن یک جمعیت به نام ازدحام وجود دارد و شامل راه‌حل‌هایی است که این راه‌حل‌ها به‌عنوان ذرات در فضای جستجو کدگذاری می‌شوند. این الگوریتم با مقداردهی اولیه‌ی ذرات به‌صورت تصادفی شروع به‌کار می‌کند، این ذرات در فضای جستجو حرکت می‌کنند و توسط به‌روزرسانی موقعیت خود بر اساس تجربه‌ی شخصی و تجربه‌ی گروه، به جستجوی راه‌حل بهینه می‌پردازند. موقعیت فعلی ذره‌ی i توسط یک بردار به صورت $x_i = (x_{i1}, x_{i2}, \dots, x_{iD})$ نشان داده می شود که D ابعاد فضای جستجو است. سرعت ذره‌ی i به‌صورت $v_i = (v_{i1}, v_{i2}, \dots, v_{iD})$ نشان داده می‌شود و با یک سرعت حداکثری از پیش تعریف‌شده (v_{max}) محدود می‌شود به‌صورتی که $v_{id}^t \in [-v_{max}, v_{max}]$. بهترین موقعیت یک ذره تاکنون به‌عنوان بهترین تجربه‌ی شخصی $(pbest)$ و بهترین موقعیت به‌دست‌آمده توسط جمع تاکنون، بهترین تجربه‌ی گروه $(gbest)$ نامیده می‌شود. به‌روزرسانی موقعیت و سرعت هر ذره توسط روابط (۱) و (۲) انجام می‌شود:

$$x_{id}^{t+1} = x_{id}^{t} + v_{id}^{t+1} \quad (۱)$$

$$v_{id}^{t+1} = w \times v_{id}^{t} + c_1 \times r_{1i} \times (p_{id} - x_{id}^{t}) + c_2 \times r_{2i} \times (p_{gd} - x_{id}^{t}) \quad (۲)$$

در معادلات بالا t نشان‌دهنده‌ی t امین تکرار و d نشان دهنده‌ی d امین بعد در فضای جستجو است $(d \in D)$، w وزن اینرسی، c_1 و c_2 ثابت‌های شتاب و r_{1i} و r_{2i} مقادیر تصادفی با توزیع یکنواخت در $[0,1]$ هستند. p_{id} و p_{gd} نشان‌دهنده‌ی $pbest$ و $gbest$ در بعد d هستند [۱۷].

۲-۲- انتخاب ویژگی ارائه‌شده توسط Xue و همکارانش در [۱۷]

همان‌گونه که در مقدمه اشاره شد Xue و همکارانش در [۱۷] از بهینه سازی ازدحام ذرات پیوسته برای انتخاب ویژگی استفاده کردند و سه راهبرد مقداردهی اولیه‌ی جدید و سه رویکرد به‌روزرسانی جدید برای بهترین تجربه‌ی شخصی و بهترین تجربه‌ی گروه ارائه دادند. ترکیب امیدبخش‌ترین راهبرد مقداردهی اولیه و رویکرد به‌روزرسانی را PSO(4-2) نامیده‌اند. یکی از سه راهبرد مقداردهی اولیه‌ی ارائه‌شده توسط آن‌ها small initialization بر اساس انتخاب روبه‌جلو است و در آن ذرات با استفاده از تعداد کمی از ویژگی‌ها مقداردهی اولیه می شوند و ترکیب ویژگی‌ها به‌صورت تصادفی انتخاب می‌شود. راهبرد دیگر، large initialization بر اساس انتخاب روبه‌عقب است و در آن ذرات با استفاده از تعداد زیادی از ویژگی‌ها مقداردهی اولیه می شوند و ترکیب ویژگی‌ها به صورت تصادفی انتخاب می شود و راهبرد ارائه شده‌ی دیگر، mixed initialization است که ترکیبی از دو حالت قبلی است که در آن بیشتر ذره‌ها با تعداد کمی ویژگی و بقیه با تعداد زیادی ویژگی، مقداردهی اولیه می‌شوند.

یکی از سه رویکرد به‌روزرسانی ارائه شده توسط Xue و همکارانش در [۱۷]، به‌روزرسانی با راهبرد کارایی رده‌بندی به‌عنوان اولویت اول است که در آن بهترین تجربه‌ی شخصی و بهترین تجربه‌ی گروه در دو حالت به‌روزرسانی می شوند. در حالت اول اگر کارایی رده‌بندی موقعیت جدید ذره از بهترین تجربه‌ی شخصی آن بهتر است به‌روزرسانی انجام می‌شود در این حالت تعداد ویژگی در نظر گرفته نمی‌شود. در حالت دوم اگر کارایی رده‌بندی موقعیت جدید ذره با بهترین تجربه‌ی شخصی آن برابر است ولی تعداد ویژگی‌های کمتری دارد به‌روزرسانی انجام می‌شود و موقعیت جدید به‌عنوان بهترین تجربه‌ی شخصی آن ذره قرار داده می‌شود. بهترین تجربه‌ی گروه به همین روش به‌روزرسانی می‌شود.

در PSO(4-2) از راهبرد mixed initialization برای مقداردهی اولیه ی ذرات استفاده می‌شود. یک بخش زیادی از جمعیت (2/3 از ذرات) با استفاده از تعداد کمی از ویژگی‌ها (حدود ۱۰ درصد کل ویژگی‌های موجود در مجموعه‌ی داده) و بخش کوچک دیگر (1/3 از ذرات) با تعداد زیادی از ویژگی ها (بیشتر از نصف تعداد ویژگی های موجود در مجموعه‌ی داده) مقداردهی اولیه می‌شوند و این برای استفاده از مزیت‌ها و جلوگیری از معایب انتخاب روبه‌جلو و انتخاب روبه‌عقب است. به‌روزرسانی بر اساس راهبرد کارایی رده‌بندی به‌عنوان اولویت اول انجام می‌گیرد. این الگوریتم نتایج امیدوارکننده‌ای از خود نشان داده است و ما از رویکردهای به‌کاررفته در PSO(4-2) در کار خود استفاده نموده و کار خود را با آن مقایسه می‌نماییم.

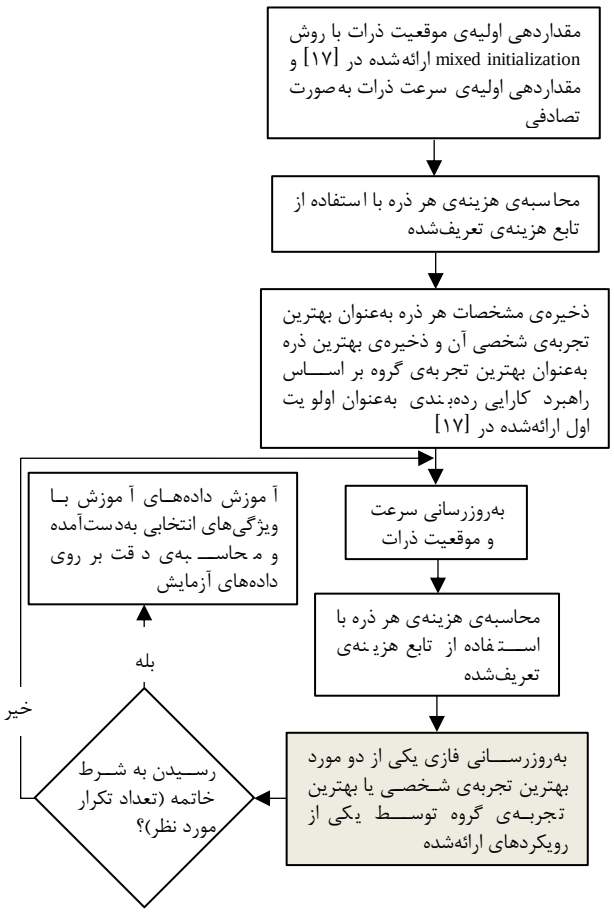
Xue و همکارانش در [۱۷] از تابع برازندگی که در رابطه‌ی (۳) نشان داده شده است استفاده کرده‌اند.

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

که در آن FP[۱۷] و FN[۱۸] و TP[۱۹] و TN[۲۰] به ترتیب نماینده‌ی مثبت کاذب، منفی کاذب، مثبت واقعی و منفی واقعی است. برای مثال، در یک مسئله‌ی رده‌بندی دودویی که شامل یک کلاس مثبت و یک کلاس منفی است، برای نمونه‌ی a، اگر کلاس a مثبت پیش‌بینی شود و برچسب کلاس واقعی a نیز مثبت باشد، TP یکی افزایش می‌یابد، اگر کلاس a مثبت پیش‌بینی شود اما برچسب کلاس واقعی a منفی باشد FP یکی افزایش می‌یابد، اگر کلاس a منفی پیش‌بینی شود و برچسب کلاس واقعی a نیز منفی باشد TN یکی افزایش می‌یابد و اگر کلاس a منفی پیش‌بینی شود اما برچسب کلاس واقعی a مثبت باشد FN یکی افزایش می‌یابد [۱۷].

۳- روش پیشنهادی

در این مقاله روشی برای انتخاب ویژگی با استفاده از بهینه‌سازی ازدحام ذرات ارائه شده است. در این الگوریتم قسمت به‌روزرسانی بهینه‌سازی ازدحام ذرات به‌صورت فازی برای انتخاب ویژگی انجام گرفته است، به این صورت که یکی از دو مورد بهترین تجربه‌ی شخصی یا بهترین تجربه‌ی گروه به‌صورت فازی به‌روزرسانی می‌شود. بنابراین با توجه به الگوریتم انتخاب ویژگی ارائه شده توسط Xue و همکارانش در [۱۷] و رویکرد ارائه‌شده در این مقاله، مراحل انتخاب ویژگی مانند شکل ۱ انجام می‌شود.



شکل ۱: نمودار گردشی روش

در ابتدا داده‌های آموزش و آزمایش به نسبت ۷۰ به ۳۰ به‌صورت تصادفی تقسیم‌بندی می‌شوند. از ۷۰ درصد داده‌های آموزش بر اساس شکل ۱ برای انتخاب ویژگی‌ها استفاده می‌شود. بعد از انتخاب ویژگی‌ها، ۷۰ در صد آموزش با ویژگی‌های به‌دست‌آمده، آموزش داده می شود و دقت بر روی ۳۰ در صد داده‌ی آزمایش محاسبه می‌گردد که از رده‌بند k-NN[۲۱] و با مقادیر پیش‌فرض برای پارامترهای آن و k=5 در این مرحله استفاده شده است.

در این مقاله مانند پژوهش ارائه‌شده توسط Xue و همکارانش در [۱۷] از بهینه سازی ازدحام ذرات پیوسته برای انتخاب ویژگی استفاده شده است که نمایش ذرات با یک آرایه‌ی n بعدی نشان داده می شود که n برابر با تعداد ویژگی‌های موجود در مجموعه‌ی داده است. اعداد موجود در هر بعد، در بازه‌ی [0,1] هستند که برای انتخاب و یا عدم انتخاب یک ویژگی بر اساس مقدار انتخاب‌شده توسط Xue و همکارانش در [۱۷] از آستانه‌ی ۰/۶ استفاده شده است. هزینه‌ی هر ذره با استفاده از تابع هزینه‌ی بخش ۳-۱ به‌دست می‌آید. بعد از اجرای الگوریتم بهینه‌سازی ازدحام ذرات به تعداد تکرار مشخص‌شده، ۷۰ درصد آموزش با ویژگی‌های انتخاب‌شده در بهترین تجربه‌ی گروه، آموزش داده می‌شود و دقت بر روی ۳۰ درصد داده‌ی آزمایش با همان ویژگی‌های انتخابی محاسبه می‌گردد که از رده‌بند k-NN و با تنظیمات گفته‌شده در پاراگراف قبل، در این مرحله استفاده شده است. تعداد نمونه‌های آزمایش که کلاس آن‌ها به‌درستی تشخیص داده شده است بر اساس درصد و به‌عنوان دقت برای یک بار اجرای برنامه در نظر گرفته می‌شود.

۴ پارامتر الگوریتم بهینه‌سازی ازدحام ذرات مانند مقادیر استفاده شده توسط Xue و همکارانش در [۱۷] در نظر گرفته شده که در جدول ۱ نشان داده شده است. حداکثر تعداد تکرار الگوریتم بهینه‌سازی ازدحام ذرات در این مقاله برای تمامی روش‌های ارائه‌شده و همچنین برای پیاده سازی روش ارائه شده توسط Xue و همکارانش [۱۷]، ۲۰۰ در نظر گرفته شده است.

جدول ۱: مقادیر پارامترها برای الگوریتم بهینه‌سازی ازدحام ذرات

نام پارامتر	مقدار
اندازه‌ی جمعیت (تعداد ذرات)	۳۰
وزن اینرسی (w)	۰/۷۲۹۸
ثابت‌های شتاب (c_1 و c_2)	۱/۴۹۶۱۸
آستانه‌ی انتخاب ویژگی	۰/۶

۳-۱- تابع هزینه‌ی مورد استفاده در الگوریتم پیشنهادی

در تابع هزینه، برای انتخاب و یا عدم انتخاب یک ویژگی بر اساس مقدار انتخاب‌شده توسط Xue و همکارانش در [۱۷] از آستانه‌ی ۰/۶ استفاده شده است. بنابراین با توجه به موقعیت ذره، عدد موجود در هر بعد اگر بیشتر از ۰/۶ باشد آن ویژگی انتخاب می‌شود، در غیر این صورت انتخاب نمی‌شود. خطای رده‌بندی برای زیرمجموعه‌ی ویژگی های انتخاب‌شده‌ی ذره، با استفاده از 10-fold cross-validation روی ۷۰

درصد داده‌های آموزش که در ابتدا تعیین شده، با ویژگی‌های انتخابی و با استفاده از k-NN با مقادیر پیش‌فرض برای پارامترها و k=5 (بر اساس تنظیمات در نظر گرفته شده برای k-NN در [۱۷])، محاسبه می شود و به‌عنوان خروجی تابع هزینه برای آن ذره در نظر گرفته می‌شود.

۳-۲- رویکردهای ارائه‌شده برای به‌روزرسانی فازی

رویکرد ارائه‌شده در این مقاله مربوط به به‌روزرسانی فازی یکی از دو مورد بهترین تجربه‌ی شخصی یا بهترین تجربه‌ی گروه است که شامل سه مورد محاسبه‌ی نمودارهای درجه تعلق، تعریف قوانین فازی و چگونگی به‌روزرسانی است. ۸ حالت مختلف در این کار ارائه شده است که سه مورد ذکرشده برای هر حالت در ادامه آورده شده است.

۳-۲-۱- محاسبه‌ی نمودارهای درجه تعلق

برای تمام حالات پیشنهادی، دو ورودی اختلاف هزینه (CostDifference) و اختلاف تعداد ویژگی‌ها (FeatureCountDifference) و یک خروجی (Value) برای سیستم فازی در نظر می‌گیریم. همان‌گونه که گفته شد ۷۰ در صد کل نمونه‌ها به‌صورت تصادفی برای آموزش و بقیه برای آزمایش انتخاب می‌شود. توسط داده‌های آموزش، انتخاب ویژگی انجام می‌شود. در تابع هزینه خطای هر زیرمجموعه از ویژگی ها توسط k-NN و 10-fold cross-validation روی ۷۰ در صد داده‌های آموزش محاسبه می شود. بنابراین در نمودار های درجه تعلق، حداقل و حداکثر اختلاف هزینه‌ی دو زیرمجموعه‌ی ویژگی را به‌صورت زیر محاسبه می‌کنیم که به‌عنوان ابتدا و انتهای نمودار ورودی اختلاف هزینه در نظر گرفته می‌شود.

- **محاسبه‌ی حداکثر و حداقل بازه‌ی محور افقی در نمودار درجه تعلق ورودی اول (اختلاف هزینه)**

برای محاسبه‌ی حداکثر برای محور افقی ورودی اختلاف هزینه (max_1)، مانند رابطه‌ی (۴)، ۷۰ درصد تعداد کل نمونه‌ها محاسبه می‌شود (تعداد داده های آموزش) و به دلیل این که از 10-fold cross-validation در محاسبه‌ی هزینه‌ی زیرمجموعه‌ی ویژگی‌ها استفاده شده است تقسیم بر ۱۰ می‌شود.

$$max_1 = \frac{70\%\ \text{of the total number of dataset instances}}{10} \quad (۴)$$

محاسبه‌ی حداقل برای محور افقی ورودی اختلاف هزینه (min_1) به صورت رابطه‌ی (۵) است.

$$min_1 = -max_1 \quad (۵)$$

- **محاسبه‌ی حداکثر و حداقل بازه‌ی محور افقی در نمودار درجه تعلق ورودی دوم (اختلاف تعداد ویژگی‌ها)**

حداکثر اختلاف تعداد ویژگی‌های انتخاب‌شده (max_2) را برابر تعداد کل ویژگی‌های مجموعه‌ی داده در نظر می‌گیریم که در رابطه‌ی (۶) نشان داده شده است.

$$max_2 = \text{total number of dataset features} \quad (۶)$$

حداقل اختلاف تعداد ویژگی‌های انتخاب شده (min_2) به صورت رابطه‌ی (۷) محاسبه می‌شود.

$$min_2 = -max_2 \quad (۷)$$

به‌عنوان مثال برای مجموعه داده‌ی Musk1 حداکثر و حداقل بازه ی محور افقی ورودی اول در شکل ۲ و حداکثر و حداقل بازه‌ی محور افقی ورودی دوم در شکل ۳ مشاهده می‌شود.

در ادامه، رویکردهای پیشنهادی مختلف برای محاسبه‌ی نمودارهای درجه تعلق، قوانین استفاده‌شده و چگونگی به‌روزرسانی در هر حالت پیشنهادی آورده شده است.

- **محاسبه‌ی نمودارهای درجه تعلق ورودی‌ها**

محاسبه‌ی نمودارهای درجه تعلق برای ورودی اختلاف هزینه و ورودی تعداد ویژگی‌ها در هر کدام از ۸ حالت در ادامه آورده شده است.

❖ حالت ۱

نمودار درجه تعلق برای ورودی اختلاف هزینه:

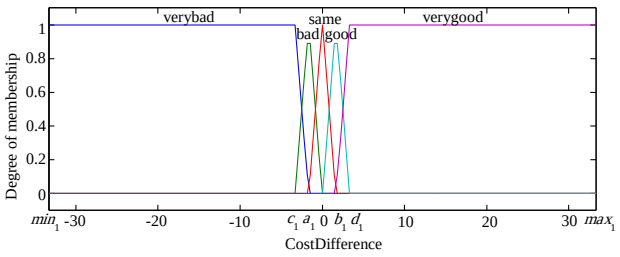
نمودار اختلاف هزینه در حالت ۱ شامل پنج مجموعه‌ی verybad، bad، same، good و verygood است و متغیرهای d_1، b_1، a_1 و c_1 روی محور افقی بر اساس روابط (۸)، (۹)، (۱۰) و (۱۱) محاسبه می‌شود. شکل ۲ این نمودار را برای مجموعه داده‌ی Musk1 نشان می‌دهد.

$$d_1 = (max_1 \times 10)/100 \quad (۸)$$

$$b_1 = (max_1 \times 5)/100 \quad (۹)$$

$$a_1 = -b_1 \quad (۱۰)$$

$$c_1 = -d_1 \quad (۱۱)$$



شکل ۲: نمودار اختلاف هزینه برای مجموعه داده‌ی Musk1 در حالت ۱

نمودار درجه تعلق برای ورودی اختلاف تعداد ویژگی‌ها:

نمودار اختلاف تعداد ویژگی ها در حالت ۱ شامل پنج مجموعه‌ی verybad، bad، same، good و verygood است و متغیرهای d_2، b_2، a_2 و c_2 روی محور افقی بر اساس روابط (۱۲)، (۱۳)، (۱۴) و (۱۵) محاسبه

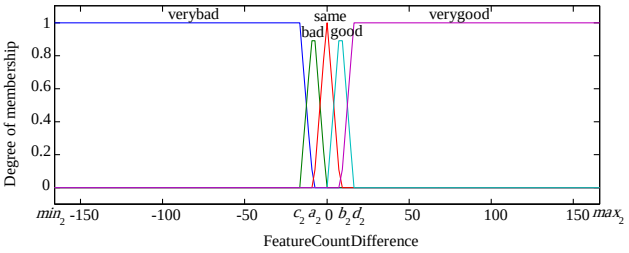
می‌شود. شکل ۳ این نمودار را برای مجموعه داده‌ی Musk1 نشان می‌دهد.

$$d_2 = (max_2 \times 10)/100 \quad (۱۲)$$

$$b_2 = (max_2 \times 5)/100 \quad (۱۳)$$

$$a_2 = -b_2 \quad (۱۴)$$

$$c_2 = -d_2 \quad (۱۵)$$



شکل ۳: نمودار اختلاف تعداد ویژگی‌ها برای مجموعه داده‌ی Musk1 در حالت ۱

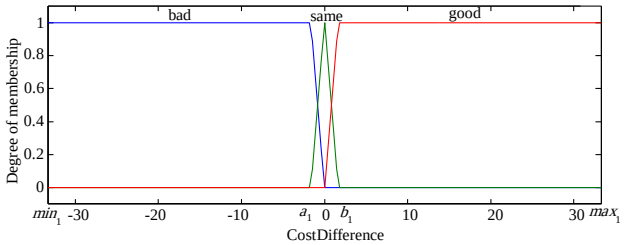
❖ حالت ۲

نمودار درجه تعلق برای ورودی اختلاف هزینه:

نمودار اختلاف هزینه در حالت ۲ شامل سه مجموعه‌ی bad، same و good است و دو متغیر b_1 و a_1 روی محور افقی بر اساس روابط (۱۶) و (۱۷) محاسبه می‌شود. شکل ۴ این نمودار را برای مجموعه داده‌ی Musk1 نشان می‌دهد.

$$b_1 = (max_1 \times 5)/100 \quad (۱۶)$$

$$a_1 = -b_1 \quad (۱۷)$$



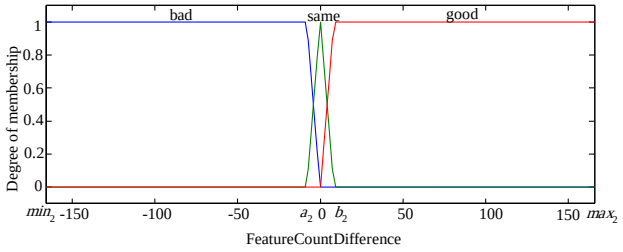
شکل ۴: نمودار اختلاف هزینه برای مجموعه داده‌ی Musk1 در حالت ۲

نمودار درجه تعلق برای ورودی اختلاف تعداد ویژگی‌ها:

نمودار اختلاف تعداد ویژگی‌ها در حالت ۲ شامل سه مجموعه‌ی bad، same و good است و دو متغیر b_2 و a_2 روی محور افقی بر اساس روابط (۱۸) و (۱۹) محاسبه می‌شود. شکل ۵ این نمودار را برای مجموعه داده‌ی Musk1 نشان می‌دهد.

$$b_2 = (max_2 \times 5)/100 \quad (۱۸)$$

$$a_2 = -b_2 \quad (۱۹)$$



شکل ۵: نمودار اختلاف تعداد ویژگی‌ها برای مجموعه داده‌ی Musk1 در حالت ۲

❖ حالات دیگر

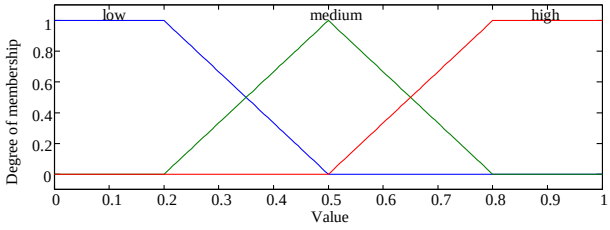
در بقیه حالات، دو ورودی اختلاف هزینه و اختلاف تعداد ویژگی ها شامل سه مجموعه‌ی bad، same و good ا ست که طبق روابط جدول ۲ محاسبه می‌شوند.

جدول ۲: روابط مربوط به دو ورودی سیستم فازی در حالات دیگر

نام حالت	محاسبه‌ی b_1 و a_1 برای ورودی اختلاف هزینه	محاسبه‌ی b_2 و a_2 برای ورودی اختلاف تعداد ویژگی‌ها
حالت ۳	$b_1 = (max_1 \times 2)/100$ $a_1 = -b_1$	$b_2 = (max_2 \times 2)/100$ $a_2 = -b_2$
حالت ۴	$b_1 = (max_1 \times 10)/100$ $a_1 = -b_1$	$b_2 = (max_2 \times 10)/100$ $a_2 = -b_2$
حالت ۵	$b_1 = (max_1 \times 2)/100$ $a_1 = -(max_1 \times 5)/100$	$b_2 = (max_2 \times 2)/100$ $a_2 = -(max_2 \times 5)/100$
حالت ۶	$b_1 = (max_1 \times 5)/100$ $a_1 = -(max_1 \times 2)/100$	$b_2 = (max_2 \times 5)/100$ $a_2 = -(max_2 \times 2)/100$
حالت ۷ و ۸	نمودارهای درجه تعلق برای دو ورودی اختلاف هزینه و اختلاف تعداد ویژگی‌ها در حالت ۷ و ۸ مانند حالت ۲ است.	

• محاسبه‌ی نمودار درجه تعلق خروجی

نمودار درجه تعلق برای خروجی در تمام حالات به‌صورت شکل ۶ در نظر گرفته شده است.



شکل ۶: نمودار درجه تعلق برای خروجی در تمامی حالات

۳-۲-۲- قوانین فازی حالات مختلف

❖ حالت ۱

1. If (CostDifference is verygood) then (Value is high) (1)

2. If (CostDifference is good) and (FeatureCountDifference is not verybad) then (Value is high) (1)

3. If (CostDifference is good) and (FeatureCountDifference is verybad) then (Value is medium) (1)

4. If (CostDifference is same) and (FeatureCountDifference is same) then (Value is medium) (1)

5. If (CostDifference is bad) and (FeatureCountDifference is verygood) then (Value is medium) (1)

6. If (CostDifference is bad) and (FeatureCountDifference is not verygood) then (Value is low) (1)

7. If (CostDifference is verybad) then (Value is low) (1)

❖ سایر حالات

1. If (CostDifference is good) then (Value is high) (1)

2. If (CostDifference is same) and (FeatureCountDifference is good) then (Value is high) (1)

3. If (CostDifference is same) and (FeatureCountDifference is same) then (Value is medium) (1)

4. If (CostDifference is same) and (FeatureCountDifference is bad) then (Value is low) (1)

5. If (CostDifference is bad) then (Value is low) (1)

۳-۲-۳- نحوه‌ی به‌روزرسانی در هر حالت

❖ ۶ حالت اول

در حالت‌های ۱، ۲، ۳، ۴، ۵ و ۶ به‌روز سانی بهترین تجربه‌ی شخصی به‌صورت فازی انجام می‌گیرد. ورودی‌های اختلاف هزینه ($Input_{1i}$) و اختلاف تعداد ویژگی‌ها ($Input_{2i}$) بر اساس رابطه‌های (۲۰) و (۲۱) محاسبه می‌شوند و به سیستم فازی داده می‌شوند.

$$Input_{1i} = pbestCost_i - pCost_i \qquad (۲۰)$$

در رابطه‌ی (۲۰)، $pbestCost_i$ نشان‌دهنده‌ی هزینه‌ی بهترین تجربه‌ی شخصی ذره‌ی i و $pCost_i$ نشان‌دهنده‌ی هزینه‌ی موقعیت جاری ذره‌ی i است. در صورت این‌که هزینه‌ی محاسبه‌شده برای موقعیت فعلی ذره ی i ($pCost_i$) از هزینه‌ی بهترین تجربه‌ی شخصی آن ($pbestCost_i$) کمتر (بهتر) باشد اختلاف هزینه‌ی ایجاد شده ($Input_{1i}$) عددی مثبت و در غیر این صورت عددی منفی خواهد شد و این عدد به‌عنوان ورودی اول به سیستم فازی داده می‌شود.

$$Input_{2i} = pbestNF_i - pNF_i \qquad (۲۱)$$

در رابطه‌ی (۲۱)، $pbestNF_i$ نشان‌دهنده‌ی تعداد ویژگی‌های بهترین تجربه‌ی شخصی ذره‌ی i و pNF_i نشان‌دهنده‌ی تعداد ویژگی‌های موقعیت جاری ذره‌ی i است. در صورت این‌که تعداد ویژگی‌های انتخاب شده توسط موقعیت فعلی ذره‌ی i (pNF_i)، کمتر (بهتر) از تعداد ویژگی‌های انتخاب‌شده در بهترین تجربه‌ی شخصی آن ($pbestNF_i$) باشد اختلاف تعداد ویژگی‌ها ($Input_{2i}$) عددی مثبت و در غیر این صورت منفی خواهد شد و این عدد به‌عنوان ورودی دوم به سیستم فازی داده می‌شود. اگر خروجی سیستم فازی بزرگ‌تر یا مساوی با ۰/۵ باشد به‌روزرسانی انجام می‌شود در غیر این صورت بهترین تجربه‌ی شخصی ذره‌ی i در این تکرار به‌روزرسانی نمی‌شود.

به‌روزرسانی بهترین تجربه‌ی گروه در این ۶ حالت بر اساس الگوریتم PSO(4-2) ارائه‌شده توسط Xue و همکارانش در [۱۷] با راهبرد کارایی رده‌بندی به‌عنوان اولویت اول انجام می‌گیرد.

❖ حالت ۷

در حالت ۷ به‌روزرسانی بهترین تجربه‌ی شخصی بر اساس الگوریتم PSO(4-2) ارائه‌شده توسط Xue و همکارانش در [۱۷] با راهبرد کارایی رده‌بندی به‌عنوان اولویت اول انجام می‌گیرد. به‌روز رسانی بهترین تجربه ی گروه، ترکیبی از به‌روزرسانی بر اساس الگوریتم PSO(4-2) و رویکرد فازی است. همان‌طور که در الگوریتم ۱ مشاهده می‌شود از تکرار اول و تا قبل از تکرار ۱۰۰ به‌روز سانی بهترین تجربه‌ی گروه، به صورت فازی انجام می‌شود.

الگوریتم ۱: شبه‌کد به‌روزرسانی حالت ۷

```
for iteration=1:Muximum of Iterations
  for i=1:Population Size
    Update the Velocity of Particle_i ;
    Update the Position of Particle_i ;
    Update the pbest_i according to classification performance as the first
    priority presented at [17] ;
    if iteration < 100
      if gbestCost < Cost of globalMinimum
        globalMinimum = gbest ;
      end
      if gbestCost is equal to Cost of globalMinimum and
        gbestNF < the number of globalMinimum features
        globalMinimum = gbest ;
      end
      Input_1i = gbestCost – pbestCost_i ;
      Input_2i = gbestNF – pbestNF_i ;
      if FuzzyOutput > 0.3
        gbest = pbest_i ;
      end
    else
      if iteration is equal to 100
        gbest = globalMinimum ;
      end
      if pbestCost_i < gbestCost
        gbest = pbest_i ;
      elseif pbestCost_i is equal to gbestCost and pbestNF_i < gbestNF
        gbest = pbest_i ;
      end
    end
  end
end
```

به این ترتیب که، ابتدا قبل از به‌روزرسانی فازی، بهترین تجربه‌ی گروه با توجه به معیار کارایی رده‌بندی به‌عنوان اولویت اول ارائه‌شده توسط Xue و همکارانش در [۱۷] با متغیر *globalMinimum* مقایسه می شود و در صورت برقراری شرط در آن ذخیره می‌شود.

محاسبه‌ی ورودی‌ها برای سیستم فازی بر اساس رابطه‌های (۲۲) و (۲۳) محاسبه می‌شود:

$$Input_{1i} = gbestCost - pbestCost_i \qquad (۲۲)$$

در رابطه‌ی (۲۲)، $gbestCost$ نشان‌دهنده‌ی هزینه‌ی بهترین تجربه‌ی گروه و $pbestCost_i$ نشان‌دهنده‌ی هزینه‌ی بهترین تجربه‌ی شخصی ذره ی i است.

$$Input_{2i} = gbestNF - pbestNF_i \quad (۲۳)$$

در رابطه‌ی (۲۳)، $gbestNF$ تعداد ویژگی‌های بهترین تجربه‌ی گروه و $pbestNF_i$ تعداد ویژگی‌های بهترین تجربه‌ی شخصی ذره‌ی i را نشان می‌دهند. اگر خروجی سیستم فازی از ۰/۳ بیشتر بود بهترین تجربه‌ی گروه به‌روزرسانی می‌شود.

در تکرار ۱۰۰، مقدار $globalMinimum$ در بهترین تجربه‌ی گروه قرار داده می‌شود و در ادامه‌ی کار تا تکرار ۲۰۰، به‌روزرسانی بهترین تجربه‌ی گروه بر اساس به‌روزرسانی الگوریتم PSO(4-2) ارائه‌شده توسط Xue و همکارانش در [۱۷] با راهبرد کارایی رده‌بندی به‌عنوان اولویت اول انجام می‌گیرد.

❖ حالت ۸

روش به‌روزرسانی در این حالت مانند حالت ۷ است، با این تفاوت که در حالت ۸ خروجی سیستم فازی با ۰/۴ مقایسه می‌شود.

۴- نتایج آزمایش‌ها

۴-۱- مجموعه داده‌های مورد استفاده

در اینجا به مجموعه داده‌های مورد استفاده در این مقاله از مخزن یادگیری ماشین UCI [۱۸] اشاره می شود. این موارد از بین مجموعه داده‌های مورد استفاده توسط Xue و همکارانش در [۱۷] انتخاب شده‌اند و مشخصات آن‌ها در جدول ۳ آمده است.

۴-۲- نتایج به‌دست‌آمده از اجرای الگوریتم‌ها

اسامی الگوریتم‌های ارائه‌شده به‌صورت جدول ۴ در نظر گرفته شده است. نتایج حاصل از میانگین ۵۰ بار اجرای الگوریتم‌های ارائه شده در این مقاله و PSO(4-2) ارائه شده توسط Xue و همکارانش در [۱۷] در جدول ۵ نشان داده شده است. در ستون تعداد ویژگی، میانگین تعداد ویژگی‌ها از ۵۰ بار اجرا نشان داده شده است. در ستون دقت، میانگین دقت‌ها بر اساس درصد از ۵۰ بار اجرا نشان داده شده است و عدد داخل پرانتز بیشترین دقت به‌دست‌آمده بین اجراها است. در تمامی الگوریتم‌ها تعداد تکرار الگوریتم بهینه‌سازی ازدحام ذرات ۲۰۰ در نظر گرفته شده و مقدار پارامترها مانند مقادیر بیان شده در فصل گذشته تنظیم شده است. از تابع هزینه‌ی توضیح داده‌شده در بخش ۳-۱ برای تمامی الگوریتم‌های پیشنهادی و PSO(4-2) استفاده شده است.

جدول ۳: مجموعه داده‌های مورد استفاده از UCI [۱۸]

نام مجموعه‌ی داده	تعداد ویژگی‌ها	تعداد نمونه‌ها	تعداد کلاس‌ها
Zoo	۱۷	۱۰۱	۷
German	۲۴	۱۰۰۰	۲
Wisconsin Diagnostic Breast Cancer (WDBC)	۳۰	۵۶۹	۲
Ionosphere	۳۴	۳۵۱	۲
Sonar	۶۰	۲۰۸	۲
Musk (Version 1) (Musk1)	۱۶۶	۴۷۶	۲

جدول ۴: اسامی الگوریتم‌های ارائه‌شده

نام حالت	نام الگوریتم	نام حالت	نام الگوریتم
حالت ۱	PbestUpdate1	حالت ۵	PbestUpdate5
حالت ۲	PbestUpdate2	حالت ۶	PbestUpdate6
حالت ۳	PbestUpdate3	حالت ۷	GbestUpdate1
حالت ۴	PbestUpdate4	حالت ۸	GbestUpdate2

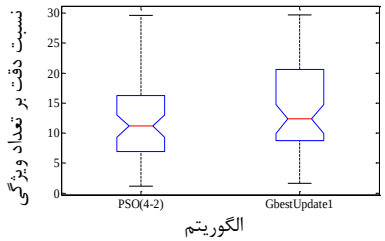
جدول ۵: نتایج به‌دست‌آمده از اجرای الگوریتم‌ها

نام مجموعه‌ی داده	الگوریتم	تعداد ویژگی (میانگین ۵۰ بار)	دقت (٪) (میانگین ۵۰ بار)	نام مجموعه‌ی داده	الگوریتم	تعداد ویژگی (میانگین ۵۰ بار)	دقت (٪) (میانگین ۵۰ بار)
Zoo	PSO(4-2)	۶/۲۸	۸۷/۵۳۳۳ (۱۰۰)	German	PSO(4-2)	۸/۱۲	۷۰/۷۵۳۳ (۷۷)
	PbestUpdate1	۶/۶۸	۸۷/۲ (۹۶/۶۶۶۷)		PbestUpdate1	۸/۶۸	۷۰/۸۱۳۳ (۷۵/۳۳۳۳)
	PbestUpdate2	۶/۱۲	۸۸/۱۳۳۳ (۱۰۰)		PbestUpdate2	۷/۲	۷۰/۲۱۳۳ (۷۴)
	PbestUpdate3	۶/۱۴	۸۶/۸ (۹۶/۶۶۶۷)		PbestUpdate3	۷/۲۲	۶۸/۹۹۳۳ (۷۵)
	PbestUpdate4	۵/۲۴	۸۷/۶۶۶۷ (۱۰۰)		PbestUpdate4	۶/۸۴	۷۰/۱۰۶۷ (۷۶)
	PbestUpdate5	۶/۰۸	۸۶ (۹۶/۶۶۶۷)		PbestUpdate5	۷/۶	۶۹/۶۸۶۷ (۷۵/۶۶۶۷)
	PbestUpdate6	۶/۷۲	۸۸/۲ (۹۶/۶۶۶۷)		PbestUpdate6	۵/۸۲	۷۰/۴۱۳۳ (۷۶/۶۶۶۷)
	GbestUpdate1	۵/۹۴	۸۷/۲ (۹۶/۶۶۶۷)		GbestUpdate1	۷/۸۸	**۷۱/۱۹۳۳** (۷۷)
	GbestUpdate2	۵/۶۲	**۸۹/۴۶۶۷** (۱۰۰)		GbestUpdate2	۶/۴۸	۷۰/۳۵۳۳ (۷۶/۶۶۶۷)
WDBC	PSO(4-2)	۵/۳۲	**۹۴/۰۴۶۸** (۹۷/۰۷۶۰)	Ionosphere	PSO(4-2)	۳/۱۸	۸۷/۸۶۶۷ (۹۷/۱۴۲۹)
	PbestUpdate1	۷/۱۲	۹۳/۷۵۴۴ (۹۸/۲۴۵۶)		PbestUpdate1	۳/۰۶	۸۷/۷۹۰۵ (۹۴/۲۸۵۷)
	PbestUpdate2	۵/۱۶	۹۳/۲۱۶۴ (۹۶/۴۹۱۲)		PbestUpdate2	۲/۹۶	۸۸/۰۱۹۰ (۹۴/۲۸۵۷)
	PbestUpdate3	۴/۳۴	۹۳/۸۹۴۷ (۹۶/۴۹۱۲)		PbestUpdate3	۲/۸۶	۸۷/۸۶۶۷ (۹۵/۲۳۸۱)
	PbestUpdate4	۶/۱۸	۹۳/۲۸۶۵ (۹۶/۴۹۱۲)		PbestUpdate4	۲/۸۴	۸۷/۹۲۳۸ (۹۴/۲۸۵۷)
	PbestUpdate5	۴/۵۲	۹۲/۹۹۴۲ (۹۷/۰۷۶۰)		PbestUpdate5	۳/۰۶	۸۸/۰۱۹ (۹۴/۲۸۵۷)
	PbestUpdate6	۵/۴۴	۹۳/۳۵۶۷ (۹۷/۰۷۶۰)		PbestUpdate6	۲/۹۶	۸۹/۰۲۸۶ (۹۵/۲۳۸۱)

	GbestUpdate1	۴/۰۸	۹۳/۵۴۳۹ (۹۷/۶۶۰۸)		GbestUpdate1	۳/۱۶	۸۸/۴۳۸۱ (۹۴/۲۸۵۷)
	GbestUpdate2	۳/۹۴	۹۳/۴۸۵۴ (۹۷/۶۶۰۸)		GbestUpdate2	۲/۹۲	**۸۹/۱۰۴۸** (۹۶/۱۹۰۵)
Sonar	PSO(4-2)	۱۱/۱۶	۷۷/۳۵۴۸ (۸۷/۰۹۶۸)	Musk1	PSO(4-2)	۶۷/۱	۸۴/۲۷۹۷ (۹۰/۹۰۹۱)
	PbestUpdate1	۱۰/۹۴	۷۵/۶۴۵۲ (۸۷/۰۹۶۸)		PbestUpdate1	۶۷/۴۸	۸۵/۲۸۶۷ (۹۱/۶۰۸۴)
	PbestUpdate2	۹/۵۸	۷۷/۱۲۹۰ (۸۵/۴۸۳۹)		PbestUpdate2	۴۹/۹۴	۸۴/۱۶۷۸ (۹۳/۰۰۷۰)
	PbestUpdate3	۹/۰۸	۷۶/۶۱۲۹ (۹۰/۳۲۲۶)		PbestUpdate3	۶۳/۷۲	۸۴/۸۱۱۲ (۹۲/۳۰۷۷)
	PbestUpdate4	۷/۸۲	۷۵/۳۵۴۸ (۸۷/۰۹۶۸)		PbestUpdate4	۴۶/۷۶	۸۴/۴۶۱۵ (۹۰/۲۰۹۸)
	PbestUpdate5	۱۰/۳	۷۶/۹۰۳۲ (۹۱/۹۳۵۵)		PbestUpdate5	۵۳/۱۲	۸۵/۱۰۴۹ (۹۰/۹۰۹۱)
	PbestUpdate6	۹/۹	۷۵/۵۱۶۱ (۸۸/۷۰۹۷)		PbestUpdate6	۵۵/۷۶	۸۴/۰۸۳۹ (۹۰/۲۰۹۸)
	GbestUpdate1	۷/۶۸	۷۶/۱۶۱۳ (۸۸/۷۰۹۷)		GbestUpdate1	۴۷/۵۴	**۸۵/۶۳۶۴** (۹۳/۷۰۶۳)
	GbestUpdate2	۷/۱۴	**۷۷/۴۸۳۹** (۸۸/۷۰۹۷)		GbestUpdate2	۲۹/۸۶	۸۴/۸۱۱۲ (۹۳/۰۰۷۰)

۴-۳- آزمون آنوا

آزمون آنوا بر روی الگوریتم PSO(4-2) ارا ئه‌شـــده توســـط Xue و همکارانش در [۱۷] و الگوریتم GbestUpdate1 پیشـــنهادی ما، به‌کار برده شده است. برای هر الگوریتم و بر روی هر یک از مجموعه داده‌های Zoo، German، WDBC، Ionosphere، Sonar و Musk1، ۱۰ نتیجه (دقت و تعداد ویژگی) به‌دست می‌آید که هر کدام میانگین ۵۰ بار اجرا است. برای هر کدام از نتایج، دقت به‌دست‌آمده از میانگین ۵۰ بار اجرا بر تعداد ویژگی به‌دست‌آمده از میانگین ۵۰ بار اجرا تقسیم می شود و توســـط این معیار و با در نظر گرفتن مقدار ۰/۲۸۳۱ برای آلفا این دو الگوریتم از هم مجزا شده‌اند. نمودار به‌دست‌آمده از آزمون آنوا در شکل ۷ نشان داده شده است.



شکل ۷: نتیجه‌ی آزمون آنوا بر اساس معیار نسبت دقت بر تعداد ویژگی

۴-۴- آزمایش بر روی مجموعه داده‌های نویزی

از awgn برای نویزی کردن مجموعه داده‌های Zoo، German، WDBC، Ionosphere، Sonar و Musk1 در متلب اســـتفاده شـــده اســـت. به این‌صورت که ستون کلاس جدا شده و بقیه داده‌ها به عنوان ورودی به دستور awgn به صورت y=awgn(x,20,'measured') داده شده است و ستون کلاس بدون تغییر اضافه گردیده است. نتیجه‌ی اجرای الگوریتم PSO(4-2) ارائه‌شـــده توســـط Xue و همکارانش در [۱۷] و الگوریتم پیشـــنهادی GbestUpdate1 بر روی مجموعه داده‌های نویزی در جدول ۶ آمده است.

۴-۵- آزمایش بر روی مجموعه داده‌های بیان ژن

از چندین مجموعه داده‌ی بیان ژن (گرفته‌شده از http://www.gems-system.org) [۱۹] برای آزمایش‌های بیشتر استفاده شده که مشخصات آن‌ها در جدول ۷ آمده است. نتایج به‌دست‌آمده از میانگین ۵۰ بار اجرای الگوریتم PSO(4-2) ارائه‌شده توسط Xue و همکارانش در [۱۷] و دو الگوریتم پیشنهادی ما به نام‌های GbestUpdate1 و GbestUpdate2 بر روی این مجموعه داده‌ها در جدول ۸ ارائه شده است.

جدول ۶: نتایج اجرای دو الگوریتم بر روی مجموعه داده‌های نویزی

نام مجموعه‌ی داده	الگوریتم	تعداد ویژگی (میانگین ۵۰ بار)	دقت (٪) (میانگین ۵۰ بار)
Zoo_Noisy	PSO(4-2)	۶/۲	۴۳/۹۳۳۳ (۶۳/۳۳۳۳)
	GbestUpdate1	۵/۳	۴۳/۰۰۰۰ (۵۶/۶۶۶۷)
German_Noisy	PSO(4-2)	۱۰/۷۸	۶۶/۲۶۶۷ (۷۱)
	GbestUpdate1	۹/۶۸	۶۶/۳۴ (۷۲/۳۳۳۳)
WDBC_Noisy	PSO(4-2)	۱۱/۶۸	۹۱/۴۰۳۵ (۹۵/۹۰۶۴)
	GbestUpdate1	۵/۷۶	۹۰/۹۹۴۲ (۹۴/۱۵۲۰)
Ionosphere_Noisy	PSO(4-2)	۳/۳۸	۸۷/۹۴۲۹ (۹۵/۲۳۸۱)
	GbestUpdate1	۳/۲۸	۸۷/۹۶۱۹ (۹۴/۲۸۵۷)
Sonar_Noisy	PSO(4-2)	۱۱/۹۸	۷۶/۵۴۸۴ (۸۷/۰۹۶۸)
	GbestUpdate1	۹/۱۸	۷۶/۲۹۰۳ (۸۸/۷۰۹۷)
Musk1_Noisy	PSO(4-2)	۶۷/۷۶	۸۴/۹۷۹۰ (۹۰/۹۰۹۱)
	GbestUpdate1	۴۹/۷	۸۴/۴۰۵۶ (۹۳/۷۰۶۳)

جدول ۷: مجموعه داده‌های مورد استفاده از [۱۹]

نام مجموعه‌ی داده	تعداد ویژگی‌ها	تعداد نمونه‌ها	تعداد کلاس‌ها
SRBCT	۲۳۰۸	۸۳	۴
Leukemia1	۵۳۲۷	۷۲	۳
DLBCL	۵۴۶۹	۷۷	۲
9_Tumors	۵۷۲۶	۶۰	۹
Brain_Tumor1	۵۹۲۰	۹۰	۵
Brain_Tumor2	۱۰۳۶۷	۵۰	۴
Leukemia2	۱۱۲۲۵	۷۲	۳
Lung_Cancer	۱۲۶۰۰	۲۰۳	۵
14_Tumors	۱۵۰۰۹	۳۰۸	۲۶

۴-۶- ارزیابی نتایج

جدول ۹ نتایج به‌دســـت‌آمده از میانگین ۵۰ بار اجرای الگوریتم PSO(4-2) ارائه‌شـــده توســـط Xue و همکارانش در [۱۷] به همراه

بالاترین دقت (میانگین ۵۰ بار اجرا) به‌دست‌آمده از روش‌های ارائه‌شده در این مقاله را در شش مجموعه داده‌ی UCI [۱۸] نشان می‌دهد. همان‌گونه که مشاهده می شود روش‌های ارائه شده در بیشتر مجموعه داده‌ها به دقت بالاتر با تعداد ویژگی کمتر از الگوریتم PSO(4-2) دست یافته‌اند. فقط در مجموعه داده‌ی WDBC الگوریتم PSO(4-2) دقت بالاتری به‌دست آورده است که در مقایسه با بهترین دقت به‌دست‌آمده از روش‌های ارائه‌شده، میانگین تعداد ویژگی آن بیشتر است. با توجه به جدول ۹ به‌طورکلی بالاترین میانگین دقت به‌دست‌آمده در روش‌های ارائه‌شده در بیشتر مجموعه داده‌های مورد استفاده از UCI [۱۸]، توسط دو الگوریتم GbestUpdate1 و GbestUpdate2 به‌دست آمده است.

جدول ۸: نتایج چندین اجرا بر روی داده‌های بیان ژن

نام مجموعه‌ی داده	الگوریتم	تعداد ویژگی (میانگین ۵۰ بار)	دقت (٪) (میانگین ۵۰ بار)
SRBCT	PSO(4-2)	۱۹۶/۰۲	**۹۳/۴۴۰۰** (۱۰۰)
	GbestUpdate1	۷۹/۷	۹۰/۴۰۰۰ (۱۰۰)
	GbestUpdate2	۱۱۲/۵	۹۳/۱۲۰۰ (۱۰۰)
Leukemia1	PSO(4-2)	۴۲۹/۷۸	۸۵/۹۰۹۱ (۱۰۰)
	GbestUpdate1	۲۰۱/۳۴	**۸۷/۰۰۰۰** (۱۰۰)
	GbestUpdate2	۳۰۳/۰۸	۸۶/۰۰۰۰ (۱۰۰)
DLBCL	PSO(4-2)	۳۱۹/۲۸	۸۹/۱۳۰۴ (۱۰۰)
	GbestUpdate1	۱۲۶/۷	۸۹/۵۶۵۲ (۱۰۰)
	GbestUpdate2	۱۹۶/۲	**۹۰/۲۶۰۹** (۱۰۰)
9_Tumors	PSO(4-2)	۱۰۱۷/۹	۳۹/۳۳۳۳ (۶۱/۱۱۱۱)
	GbestUpdate1	۶۳۰/۷	**۴۱/۸۸۸۹** (۶۱/۱۱۱۱)
	GbestUpdate2	۶۳۵/۶	۳۷/۴۴۴۴ (۶۱/۱۱۱۱)
Brain_Tumor1	PSO(4-2)	۴۴۸/۰۲	**۸۱/۶۲۹۶** (۹۶/۲۹۶۳)
	GbestUpdate1	۲۰۱/۲۲	۷۹/۸۵۱۹ (۹۶/۲۹۶۳)
	GbestUpdate2	۱۷۹/۵	۸۱/۵۵۵۶ (۱۰۰)
Brain_Tumor2	PSO(4-2)	۱۰۳۸/۹	**۷۰/۵۳۳۳** (۹۳/۳۳۳۳)
	GbestUpdate1	۴۷۴/۱۴	۶۸/۱۳۳۳ (۸۶/۶۶۶۷)
	GbestUpdate2	۶۰۶/۵	۶۸/۰۰۰۰ (۹۳/۳۳۳۳)
Leukemia2	PSO(4-2)	۶۳۷/۳۴	**۹۰/۳۶۳۶** (۱۰۰)
	GbestUpdate1	۳۳۸/۵	۸۹/۴۵۴۵ (۱۰۰)
	GbestUpdate2	۴۱۶/۳	۸۹/۷۲۷۳ (۱۰۰)
Lung_Cancer	PSO(4-2)	۱۵۰۸/۳۸	**۹۱/۰۸۲۰** (۹۸/۳۶۰۷)
	GbestUpdate1	۷۱۱	۹۰/۲۲۹۵ (۹۶/۷۲۱۳)
	GbestUpdate2	۴۶۱/۶۴	۹۱/۰۴۹۲ (۹۸/۳۶۰۷)
14_Tumors	PSO(4-2)	۴۹۱۷/۵۴	**۴۸/۱۵۲۲** (۵۵/۴۳۴۸)
	GbestUpdate1	۲۳۹۵/۸۴	۴۷/۹۵۶۵ (۵۸/۶۹۵۷)
	GbestUpdate2	۱۵۵۴/۷۶	۴۷/۱۰۸۷ (۵۷/۶۰۸۷)

با توجه به نتایج ارائه شده از میانگین ۵۰ بار اجرا در جدول ۸ برای مجموعه داده های بیان ژن مورد استفاده از [۱۹]، الگوریتم های پیشنهادی GbestUpdate1 و GbestUpdate2 در دو مجموعه داده‌ی Leukemia1 و DLBCL به دقت بالاتر با تعداد ویژگی کمتر از الگوریتم PSO(4-2) ارائه شده توسط Xue و همکارانش در [۱۷] دست یافته‌اند. در مجموعه داده‌ی 9_Tumors الگوریتم GbestUpdate1 دقت بالاتر با تعداد ویژگی کمتر از الگوریتم PSO(4-2) به‌دست آورده است. در بقیه‌ی مجموعه داده های بیان ژن مورد استفاده از [۱۹]، الگوریتم PSO(4-2) به دقت بالاتری البته با تعداد ویژگی بیشتر از دو الگوریتم پیشنهادی ما دست یافته است که در بیشتر این موارد نیز دقت به‌دست‌آمده توسط یکی از دو الگوریتم پیشنهادی ما اختلاف کمی با دقت الگوریتم PSO(4-2) داشته است.

جدول ۹: مقایسه‌ی نتایج به‌دست‌آمده از PSO(4-2) با بهترین دقت روش‌های ارائه‌شده بر روی شش مجموعه داده‌ی UCI [۱۸]

نام مجموعه‌ی داده	الگوریتم	تعداد ویژگی (میانگین ۵۰ بار)	دقت (٪) (میانگین ۵۰ بار)
Zoo	PSO(4-2)	۶/۲۸	۸۷/۵۳۳۳ (۱۰۰)
	GbestUpdate2	**۵/۶۲**	**۸۹/۴۶۶۷** (۱۰۰)
German	PSO(4-2)	۸/۱۲	۷۰/۷۵۳۳ (۷۷)
	GbestUpdate1	**۷/۸۸**	**۷۱/۱۹۳۳** (۷۷)
WDBC	PSO(4-2)	۵/۳۲	**۹۴/۰۴۶۸** (۹۷/۰۷۶۰)
	PbestUpdate3	**۴/۳۴**	۹۳/۸۹۴۷ (۹۶/۴۹۱۲)
Ionosphere	PSO(4-2)	۳/۱۸	۸۷/۸۶۶۷ (۹۷/۱۴۲۹)
	GbestUpdate2	**۲/۹۲**	**۸۹/۱۰۴۸** (۹۶/۱۹۰۵)
Sonar	PSO(4-2)	۱۱/۱۶	۷۷/۳۵۴۸ (۸۷/۰۹۶۸)
	GbestUpdate2	**۷/۱۴**	**۷۷/۴۸۳۹** (۸۸/۷۰۹۷)
Musk1	PSO(4-2)	۶۷/۱	۸۴/۲۷۹۷ (۹۰/۹۰۹۱)
	GbestUpdate1	**۴۷/۵۴**	**۸۵/۶۳۶۴** (۹۳/۷۰۶۳)

۵- نتیجه‌گیری

انتخاب ویژگی یکی از مسائل مهم در رده‌بندی است. پژوهش‌های بسیاری در انتخاب ویژگی مبتنی بر بهینه‌سازی ازدحام ذرات انجام شده است که در این مقاله روشی برای بهبود یکی از الگوریتم‌های معرفی شده توسط نویسندگان در این زمینه ارائه شد. برای این منظور از به‌روزرسانی فازی برای یکی از دو مورد بهترین تجربه‌ی شخصی یا بهترین تجربه‌ی گروه استفاده گردید. دو ورودی اختلاف هزینه و اختلاف تعداد ویژگی‌ها به‌عنوان ورودی‌های سیستم فازی در نظر گرفته شد و چندین حالت برای محاسبه‌ی درجه تعلق‌ها ارائه شد. دو حالت برای قوانین فازی و رویکردهایی برای به‌روزرسانی در این زمینه ارائه گردید. از k نزدیک‌ترین همسایه به‌عنوان رده‌بند استفاده شد. با توجه به نتایج شبیه‌سازی‌های انجام‌شده، استفاده از به‌روزرسانی فازی

پیشنهادی باعث شد نتایج مطلوبی از لحاظ دقت و تعداد ویژگی نسبت به مقاله‌ی مرجع به‌دست آید.

مراجع

[1] B. Xue, *Particle Swarm Optimisation for Feature Selection in Classification*, ph.D. Thesis, Victoria University, Wellington, 2014.

[2] R. Kohavi and D. Sommerfield, "Feature subset selection using the wrapper method: overfitting and dynamic search space topology," *KDD-95 Proceedings*, pp. 192-197, 1995.

[3] M. Dash and H. Liu, "Feature selection for classification," *Intelligent Data Analysis*, vol. 1, no. 3, pp. 131-156, 1997.

[4] G. Chandrashekar and F. Sahin, "A survey on feature selection methods," *Computers and Electrical Engineering*, vol. 40, no. 1, pp. 16-28, 2014.

[5] C. S. Yang, L. Y. Chuang, C. H. Ke and C. H. Yang, "Boolean binary particle swarm optimization for feature selection," *2008 IEEE Congress on Evolutionary Computation (IEEE World Congress on Computational Intelligence)*, pp. 2093-2098, 2008.

[6] J. Kennedy and R. Eberhart, "Particle swarm optimization," *Neural Networks, 1995. Proceedings., IEEE International Conference on*, vol. 4, pp. 1942-1948, 1995.

[7] Y. Shi and R. Eberhart, "A modified particle swarm optimizer," *Evolutionary Computation Proceedings, 1998. IEEE World Congress on Computational Intelligence., The 1998 IEEE International Conference on*, pp. 69-73, 1998.

[8] X. Wang, J. Yang, X. Teng, W. Xia and R. Jensen, "Feature selection based on rough sets and particle swarm optimization," *Pattern Recognition Letters*, vol. 28, no. 4, pp. 459-471, 2007.

[9] C. L. Huang and J. F. Dun, "A distributed PSO–SVM hybrid system with feature selection and parameter optimization," *Applied Soft Computing*, vol. 8, no. 4, pp. 1381-1391, 2008.

[10] L. Y. Chuang, H. W. Chang, C. J. Tu and C. H. Yang, "Improved binary PSO for feature selection using gene expression data," *Computational Biology and Chemistry*, vol. 32, no. 1, pp. 29-38, 2008.

[11] M. A. Esseghir, G. Goncalves and Y. Slimani, "Adaptive particle swarm optimizer for feature selection," *Intelligent Data Engineering and Automated Learning-IDEAL 2010*, pp. 226-233, 2010.

[12] L. Y. Chuang, S. W. Tsai and C. H. Yang, "Improved binary particle swarm optimization using catfish effect for feature selection," *Expert Systems with Applications*, vol. 38, no. 10, pp. 12699-12707, 2011.

[13] A. Unler, A. Murat and R. B. Chinnam, "mr^{2}PSO: A maximum relevance minimum redundancy feature selection method based on swarm intelligence for support vector machine classification," *Information Sciences*, vol. 181, no. 20, pp. 4625-4641, 2011.

[14] S. M. Vieira, L. F. Mendonça, G. J. Farinha and J. M. Sousa, "Modified binary PSO for feature selection using SVM applied to mortality prediction of septic patients," *Applied Soft Computing*, vol. 13, no. 8, pp. 3494-3504, 2013.

[15] P. Ghamisi and J. A. Benediktsson, "Feature selection based on hybridization of genetic algorithm and particle swarm optimization," *IEEE Geoscience and Remote Sensing Letters*, vol. 12, no. 2, pp. 309-313, 2015.

[16] S. Gunasundari and S. Janakiraman, "A hybrid PSO-SFS-SBS algorithm in feature selection for liver cancer data," *Power Electronics and Renewable Energy Systems*, pp. 1369-1376, 2015.

[17] B. Xue, M. Zhang and W. N. Browne, "Particle swarm optimisation for feature selection in classification: Novel initialisation and updating mechanisms," *Applied Soft Computing*, vol. 18, pp. 261-276, 2014.

[18] M. Lichman, UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Sciences, 2013.

[19] A. Statnikov, C. F. Aliferis and I. Tsamardinos, GEMS: Gene Expression Model Selector [http://www.gems-system.org], 2005.

زیرنویس‌ها

[1] redundant

[2] irrelevant

[3] Particle Swarm Optimization

[4] rough set

[5] wrapper

[6] Support Vector Machine

[7] global best

[8] filter

[9] fitness

[10] mutual information

[11] septic

[12] variability

[13] Sequential Forward Selection

[14] Sequential Backward Selection

[15] personal best

[16] University of California, Irvine

[17] False Positives

[18] False Negatives

[19] True Positives

[20] True Negatives

[21] *k*-nearest neighbor