

ارائه یک راهکار انتخاب ویژگی چند هدفه با استفاده از نظریه اطلاعات متقابل و نظریه مجموعه پارتو

مریم رحمانی نیا^۱، دانشجوی دکتری؛ پرهام مرادی^۲، استاد؛ مهدی جلیلی^۳، استاد

۱- گروه مهندسی کامپیوتر - واحد سنندج- دانشگاه آزاد اسلامی - سنندج - ایران - ma.rahmaninia@gmail.com

۲- گروه مهندسی کامپیوتر - دانشگاه کردستان - سنندج - ایران - p.moradi@uok.ac.ir

۳- دانشکده مهندسی، دانشگاه RMIT، ملبورن، استرالیا - mahdi.jalili@rmit.edu.au

چکیده: انتخاب ویژگی، فرایند انتخاب زیرمجموعه‌ای از میان مجموعه ویژگی‌های اولیه است، بطوری‌که با حذف ویژگی‌های اضافی و نامربوط دقت دسته‌بندی افزایش یابد. روش‌های انتخاب ویژگی فیلتر به دلیل پیچیدگی محاسباتی پایین، مقیاس‌پذیری از نظر ابعاد داده‌ها و استقلال از انواع دسته‌بندی از اهمیت بالایی برخوردار هستند. اما یکی از نقاط ضعف این دسته‌بندی‌ها، کمبود اطلاعات در مورد تعامل و ارتباطات بین ویژگی‌ها است که منجر به انتخاب ویژگی‌های افزونه و نامربوط می‌شود. انتخاب ویژگی‌های افزونه و نامربوط به دلیل انتخاب نامناسب تابع هدفی است که بر اساس آن میزان اهمیت و افزونگی ویژگی‌ها تخمین زده می‌شود. در این مقاله یک روش انتخاب ویژگی فیلتر غیرخطی بر اساس اطلاعات متقابل شرطی و مجموعه پارتو ارائه و به منظور نشان دادن کارایی آن، یک سری آزمایش‌ها بر روی ۱۲ مجموعه داده آموزشی پرکاربرد انجام شده است. طبق نتایج به دست آمده، الگوریتم پیشنهادی در مقایسه با تعدادی از الگوریتم‌های انتخاب ویژگی اخیر از دقت بالاتری برخوردار است.

واژه‌های کلیدی: نظریه اطلاعات، داده آموزشی با ابعاد بالا، انتخاب ویژگی، روش‌های فیلتر، مجموعه پارتو.

A Multi-Objective Feature Selection Method based on the Mutual Information Theory and Pareto Set Theory

M. Rahmaninia¹, PhD Student; P. Moradi², Professor; M. Jalili³, Professor

1- Department of Computer Engineering, Sanandaj Branch, Azad University, Sanandaj, Iran, Email: ma.rahmaninia@gmail.com

2- Department of Computer Engineering, University of Kurdistan, Sanandaj, Iran, Email: p.moradi@gmail.com

3- School of Engineering, RMIT University, Melbourne, Australia, Email: mahdi.jalili@rmit.edu.au

Abstract: Feature selection is the process of selecting a subset of features among the set of primary features, so that, by removing the redundant and irrelevant features, the accuracy of the classification increases. Because of the low computational complexity, scalability in term of data dimensions and independence of any classifier, filter selection methods are very important. But one of the weaknesses of these methods is the lack of information about the interaction and communication between the features which leads to select redundant and irrelevant features. Selection of redundant and irrelevant features is due to the inappropriate selection of an objective function which estimates the significance and redundancy of the features. In this paper, a nonlinear filter feature selection method, based on conditional mutual information and Pareto set is presented and to prove the efficiency of it a series of experiments are performed on twelve widely used datasets. According to the results, the proposed method is more accurate than a number of recently feature selection methods.

Keywords: Information theory, High dimensional data set, Feature selection, Filter methods, Pareto set.

تاریخ ارسال مقاله: ۱۳۹۷/۰۸/۰۳

تاریخ اصلاح مقاله: ۱۳۹۷/۱۲/۲۷ و ۱۳۹۸/۰۶/۲۷

تاریخ پذیرش مقاله: ۱۳۹۸/۰۷/۲۸

نام نویسنده مسئول: دکتر پرهام مرادی

آدرس نویسنده مسئول: ایران- کردستان- سنندج- دانشگاه کردستان- دانشکده مهندسی برق و کامپیوتر

۱. مقدمه

برچسب کلاس مسئله دارند یا به عبارتی دیگر هیچ تأثیری در دسته‌بندی داده‌های آموزشی ندارند. همچنین ویژگی‌های اضافی، ویژگی‌هایی هستند که بیشترین شباهت و یا ارتباط را باهمدیگر دارند و اطلاعات جدیدی در مورد مسئله اصلی ارائه نمی‌دهند. به‌طور کلی روش‌های انتخاب ویژگی را بر اساس نحوه ارزیابی زیرمجموعه‌های انتخابی از ویژگی‌ها به دو دسته پوششی و فیلتر دسته‌بندی می‌کنند. در روش‌های پوششی در ابتدا با استفاده از یک روش جستجوی از پیش تعریف شده زیرمجموعه‌ای از ویژگی‌ها انتخاب و سپس با استفاده از یک الگوریتم یادگیر یا دسته‌بند، مجموعه انتخابی از ویژگی‌ها مورد ارزیابی قرار می‌گیرد [۲۶، ۲۷]. اگرچه جواب‌هایی که توسط این دسته از الگوریتم‌ها تولید می‌شود از دقت خوبی برخوردار هستند، اما معمولاً نتایج به‌دست‌آمده وابسته به الگوریتم یادگیر استفاده شده است. همچنین این روش‌ها به دلیل استفاده از الگوریتم یادگیری در فرآیند جستجو، از سرعت پایینی برخوردارند و برای اجرا بر روی مجموعه داده‌های بزرگ مناسب نیستند.

در مقابل، روش‌های فیلتر مستقل از دسته‌بند هستند و ویژگی‌ها را بر اساس ارتباطشان با برچسب کلاس و بر اساس اصول نظریه اطلاعات ارزیابی می‌کنند. تاکنون روش‌های زیادی به‌منظور تعیین میزان ارتباط و وابستگی یک ویژگی به کلاس مسئله یا سایر ویژگی‌ها ارائه شده است از جمله ضریب همبستگی پیر سن [۲۸]، نرخ فزاینده فیشر [۲۹]، معیار پراکندگی [۳۰]، طبقه‌بندی متغیر مجزا [۳۱]، اطلاعات متقابل [۳۲]، الگوریتم رلیف [۳۳]، نظریه مجموعه‌های راف [۳۴] و تجزیه و تحلیل پوشش داده [۳۵] را می‌توان نام برد. روش‌های فیلتر به دلیل اینکه از هیچ‌گونه الگوریتم یادگیری در فرآیند جستجو استفاده نمی‌کنند، کارایی محاسباتی و مقیاس‌پذیری بالایی دارند و نتیجه به‌دست‌آمده از آن‌ها مستقل از انواع الگوریتم‌های یادگیری است. هدف اصلی این روش‌ها بیشینه‌سازی ارتباط بین ویژگی‌های انتخاب شده با برچسب کلاس و کمینه‌سازی ارتباط درونی ویژگی‌های انتخاب شده است. به‌عنوان مثال در روش‌های [۳۲] و [۳۶] از ترکیب خطی معیار افزونگی درون ویژگی‌ها و معیار ارتباط ویژگی‌ها و کلاس، به‌عنوان تابع هدف در انتخاب ویژگی استفاده می‌شود. در این روش‌ها با افزایش تعداد ویژگی‌های انتخاب شده معیار افزونگی که عبارت است از حاصل جمع میزان ارتباط ویژگی‌های انتخاب شده با یکدیگر، تحت تأثیر معیار ارتباط قرار می‌گیرد که منجر به انتخاب نامناسب ویژگی‌ها می‌شود. این مشکل تا حد زیادی توسط روش‌های دیگر مانند [۳۷] و [۲۳] با تقسیم معیار افزونگی بر تعداد ویژگی‌های انتخاب شده تا آن لحظه مرتفع شده است. علاوه بر این در بعضی روش‌ها مانند MIFS، MIFS-U، MRMR و NMIFS معیار ارتباط ویژگی‌ها و کلاس بدون در نظر گرفتن ویژگی‌های انتخاب شده قبلی محاسبه می‌شود. این در حالی است که میزان ارتباط یک ویژگی به کلاس مسئله باید علاوه بر برچسب کلاس، سایر ویژگی‌های انتخاب شده را نیز در محاسبه خود لحاظ کند. چون هرچقدر یک ویژگی با برچسب کلاس مرتبط باشد و

رشد اینترنت و تکنولوژی‌های مربوطه منجر به تولید حجم عظیمی از داده‌ها شده است که به آن کلان داده هم گفته می‌شود. یکی از راهکارهای مؤثر برای تحلیل کلان داده‌ها، استفاده از الگوریتم‌های یادگیری ماشین است. یکی از چالش‌های اصلی الگوریتم‌های یادگیری ماشین در مواجهه با کلان داده‌ها مسئله ازدحام ابعاد است [۱]. مسئله ازدحام ابعاد، به پراکندگی داده‌های آموزشی در ابعاد بالا که باعث کاهش عملکرد الگوریتم‌های یادگیری ماشین می‌شود اشاره دارد. یکی از مؤثرترین راهکارها برای حل چالش‌های مطرح شده در حوزه داده‌های حجیم، کاهش ابعاد داده‌ها است.

الگوریتم‌های کاهش بعد به دو دسته روش‌های کاهش بعد مبتنی بر استخراج ویژگی و روش‌های کاهش بعد مبتنی بر انتخاب ویژگی دسته‌بندی می‌شوند. روش‌های کاهش بعد مبتنی بر استخراج ویژگی، فضای ویژگی‌ها را به فضایی با ابعاد کوچک‌تر نگاشت می‌کنند. فضای ویژگی‌های جدید معمولاً ترکیب خطی یا غیرخطی از فضای ویژگی‌های اصلی است. از این دسته روش‌ها می‌توان به تجزیه و تحلیل مؤلفه اصلی [۲]، تجزیه و تحلیل متمایز خطی [۳]، تحلیل همبستگی کانونیک [۴]، تجزیه مقدار منفرد [۵] و مکان‌یابی خطی محلی [۶] اشاره کرد.

از طرف دیگر، روش‌های کاهش بعد مبتنی بر انتخاب ویژگی، مجموعه‌ای از ویژگی‌های مؤثر و با ارزش را انتخاب می‌کنند که موجب افزایش سرعت الگوریتم‌های یادگیری ماشین و بهبود عملکرد آن‌ها می‌شود [۷-۱۰]. انتخاب ویژگی به‌عنوان یک موضوع تحقیقاتی مهم و فعال در شناسایی الگو، یادگیری ماشین، داده‌کاوی و آمار به‌حساب می‌آید و به‌طور گسترده در بسیاری از حوزه‌ها مانند دسته‌بندی اسناد متنی [۱۱-۱۳]، شناسایی چهره [۱۴، ۱۵]، تشخیص سرطان [۱۸-۱۶]، کارهای مالی [۱۹، ۲۰] و مدیریت ارتباط با مشتری [۲۱] بکار برده شده است. انتخاب ویژگی فرایندی است که هدف آن انتخاب زیرمجموعه‌ای کوچک از ویژگی‌های کارا و مؤثر جهت کاهش ابعاد داده است. جستجوی کامل فضای ویژگی تضمین می‌کند که مناسب‌ترین ویژگی‌ها پیدا شود. اما معمولاً این کار به لحاظ هزینه محاسباتی، حتی برای مجموعه‌های داده‌ای با اندازه متوسط، غیرممکن است. بنابراین در حالت کلی مسئله انتخاب ویژگی از نوع ان پی سخت است [۲۲-۲۵].

این روش‌ها با حفظ ویژگی‌های اصلی، مفهوم اصلی ویژگی‌ها را حفظ کرده و باعث تفسیر بهتر داده‌ها می‌شود. از طرف دیگر داده‌های دنیای واقعی معمولاً دارای یک مجموعه ویژگی‌های اضافی^۲ و نامربوط^۳ هستند.

حذف این ویژگی‌ها علاوه بر این که باعث کاهش پیچیدگی محاسباتی و حافظه مورد نیاز جهت ذخیره‌سازی می‌شود، بلکه منجر به افزایش عملکرد الگوریتم‌های یادگیری می‌شود. ویژگی‌های نامربوط، به آن دسته از ویژگی‌ها اطلاق می‌شود که کمترین ارتباط را با

معیار اطلاعات متقابل برای ارزیابی میزان وابستگی دو متغیر تصادفی ارائه شود. آنتروپی یکی از معیارهای اندازه‌گیری برای محاسبه میزان عدم قطعیت یا میزان بی‌نظمی یک متغیر است. آنتروپی یک متغیر X یا بردار با مقادیر گسسته به صورت زیر تعریف می‌شود:

$$H(X) = -\sum_{x \in X} p(x) \log p(x) \quad (1)$$

در این رابطه $p(x)$ احتمال رخداد $x \in X$ را نشان می‌دهد. همچنین آنتروپی توأم دو متغیر تصادفی X و Y که لزوماً از هم مستقل نیستند به صورت زیر تعریف می‌شود:

$$H(X, Y) = -\sum_{x \in X} \sum_{y \in Y} p(x, y) \log p(x, y) \quad (2)$$

بطوریکه $p(x, y)$ به احتمال رخداد توأم دو ویژگی X و Y اشاره می‌کند. همچنین اگر یکی از دو متغیر X یا Y مشخص باشد، مقدار عدم قطعیت متغیر دیگر به صورت زیر تعریف می‌شود:

$$H(Y|X) = -\sum_{x \in X} \sum_{y \in Y} p(x, y) \log p(y|x) \quad (3)$$

در این رابطه $p(y|x)$ به احتمال شرطی اشاره دارد و به صورت $p(x|y) = p(x, y)/p(y)$ تعریف می‌شود. حال با استفاده از آنتروپی و آنتروپی توأم، آنتروپی شرطی با استفاده از رابطه زیر تعریف می‌شود:

$$H(Y|X) = H(X, Y) - H(X) \quad (4)$$

در نظریه اطلاعات، اطلاعات متقابل بین دو متغیر به مقداری از اطلاعات که بین دو متغیر به اشتراک گذاشته شده است، اشاره می‌کند که آن را می‌توان با استفاده از رابطه زیر به دست آورد:

$$I(Y|X) = -\sum_{x \in X} \sum_{y \in Y} p(x, y) \log \frac{p(x, y)}{p(x)p(y)} \quad (5)$$

رابطه (۵) همواره مثبت است. اگر اطلاعات متقابل میان دو ویژگی زیاد باشد به این معنی است که وابستگی دو متغیر زیاد است. **لم ۱.** اطلاعات متقابل با استفاده از آنتروپی معادل متغیرها به صورت زیر قابل محاسبه است:

$$\begin{aligned} I(X; Y) &= H(Y) - H(Y|X) \\ &= H(X) - H(X|Y) \\ &= H(X) + H(Y) - H(X, Y) \end{aligned} \quad (6)$$

اثبات: در مرجع [۵۱] ارائه شده است.

حال اطلاعات متقابل میان دو متغیر X و Y به شرط وجود متغیر Z اطلاعات متقابل شرطی نامیده می‌شود که به صورت زیر تعریف می‌شود:

اطلاعات زیادی در مورد کلاس فراهم کند باز ممکن است آن اطلاعات توسط دیگر ویژگی‌ها نیز کشف شده باشد و بی‌تأثیر باشد. علاوه بر این در محاسبه میزان افزونگی، علاوه بر استفاده از اطلاعات ویژگی‌های انتخاب شده قبلی، باید از اطلاعات برچسب کلاس نیز استفاده شود. چون ممکن است دو ویژگی اطلاعات یکسانی باهم به اشتراک بگذارند اما هرکدام از آن‌ها اطلاعات متفاوتی با برچسب کلاس داشته باشند. همچنین، در روش‌هایی که از ترکیب خطی معیارهای ارتباط و افزونگی استفاده می‌کنند، پیدا کردن ضریبی که بتواند تعامل مناسبی بین این دو معیار پیدا کند مشکل است و به داده‌های آموزشی بستگی دارد.

در این مقاله یک روش انتخاب ویژگی فیلتر چندهدفه برای انتخاب مجموعه‌ای از ویژگی‌های مرتبط و غیر افزونه به منظور مرتفع کردن چالش‌های مربوط به راهکارهای قبلی ارائه شده است. در راهکار پیشنهادی که بنام Conditional Joint Mutual Information Maximization (CJMIM) نام‌گذاری شده است دو تابع هدف جدید بر اساس اطلاعات متقابل شرطی ارائه شده است. تابع اول معیار ارتباط را بیشینه می‌کند و هدف تابع دوم کمینه کردن افزونگی بین ویژگی‌های انتخاب شده است. در این توابع هم از برچسب کلاس و هم اطلاعات ویژگی‌های انتخاب شده قبلی در محاسبه میزان افزونگی و ارتباط ویژگی‌ها به صورت هم‌زمان استفاده شده است. همچنین برخلاف تمام راهکارهای قبلی که از ترکیب خطی یا غیرخطی توابع هدف استفاده کرده‌اند، در این مقاله از مجموعه پارتو به منظور انتخاب ویژگی‌هایی که دارای ارتباط بیشینه و افزونگی حداقل هستند استفاده شده است. در این راهکار با استفاده از اهداف تعریف شده و همچنین با به‌کارگیری نظریه مجموعه پارتو، جواب‌های غالب و مغلوب شناسایی شده و بر این مبنای ویژگی با افزونگی کم و ارتباط بالا انتخاب می‌شود.

برای نشان دادن کارایی الگوریتم پیشنهادی، مجموعه‌ای از شبیه‌سازی‌ها بر روی ۱۲ مجموعه داده‌ای پرکاربرد انجام شده و نتایج به دست آمده از روش پیشنهادی با هفت الگوریتم انتخاب ویژگی فیلتر از جمله [۳۲] MIFS، [۳۶] MIFS_ND، [۳۷] MRMR، [۲۳] NMIFS، [۴۶] ECFS، [۴۷] infFS و [۵۲] MOEA/D مورد مقایسه قرار گرفته است. نتایج به دست آمده نشان‌دهنده برتری روش پیشنهادی نسبت به سایر روش‌های انتخاب ویژگی است. ساختار کلی این مقاله به این صورت است: در بخش دوم راهکارهای پیشین و مفاهیم مرتبط با نظریه اطلاعات شرح داده شده است. در بخش سوم الگوریتم پیشنهادی معرفی شده و در نهایت در بخش چهارم و پنجم بحث و نتیجه‌گیری کلی ارائه شده است.

۲. پیشینه تحقیق و راهکارهای پیشین

۲-۱- نظریه اطلاعات متقابل

مفهوم اطلاعات متقابل با مفهوم آنتروپی یک متغیر تصادفی که میزان اطلاعات موجود در یک متغیر تصادفی را نشان می‌دهد مرتبط است [۴۲]. به همین دلیل در این بخش سعی می‌شود تا مفاهیم مربوط به

دسته تک متغیره و چند متغیره دسته‌بندی می‌شوند. در راهکار تک متغیره فقط ارتباط ویژگی با کلاس در محاسبات لحاظ می‌شود ولی افزونگی ویژگی‌های انتخاب شده در رتبه‌بندی ویژگی‌ها در نظر گرفته نمی‌شود. الگوریتم‌هایی مانند بهره اطلاعاتی [۱۱]، ضریب همبستگی پیرسن [۲۸]، نرخ فزاینده فیشر [۲۹] و ناپایداری متقارن^۴ [۴۳، ۱۰] را می‌توان از این دسته الگوریتم‌ها نام برد. سادگی و پیچیدگی محاسباتی کم دو خصوصیت بسیار مهم این روش‌ها است. این روش‌ها وابستگی بین ویژگی‌ها را در نظر نمی‌گیرند و معمولاً فرض می‌کنند ویژگی‌ها از همدیگر مستقل هستند. برای رفع این مشکل، روش‌های چند متغیره ارائه شده‌اند که در آن‌ها علاوه بر در نظر گرفتن میزان وابستگی ویژگی‌ها به کلاس، میزان وابستگی ویژگی‌ها به همدیگر نیز در رتبه‌بندی ویژگی‌ها در نظر گرفته می‌شود. روش‌های زیرفضای تصادفی^۵ [۸]، وابستگی متقابل^۶ [۴۴]، روش انتخاب ویژگی افزونگی - ارتباط [۷] و روش حداقل افزونگی - حداکثر ارتباط [۲۳] از نوع روش‌های انتخاب ویژگی چند متغیره فیلتر هستند. در بیشتر راهکارهای فیلتر ارائه شده، از نظریه اطلاعات متقابل برای محاسبه میزان وابستگی دو ویژگی استفاده شده است. در [۳۲]، یک روش انتخاب ویژگی افزایشی بنام MIFS ارائه شده است. این روش در ابتدا بر اساس تابع هدف زیر به هر کدام از ویژگی‌هایی که هنوز انتخاب نشده‌اند یک امتیاز داده و سپس ویژگی با بیشترین امتیاز به مجموعه ویژگی‌های انتخاب شده اضافه می‌شود:

$$I(x_i; C) - \beta \sum_{x_s \in S} I(x_i, x_s) \quad (11)$$

در این فرمول k نشان‌دهنده مجموعه ویژگی‌های انتخاب شده است. انتخاب ویژگی‌ها با استفاده از تابع هدف بالا تا زمانی که تعداد مشخصی از ویژگی‌ها انتخاب شوند ادامه می‌یابد. در مقاله [۳۶] روشی به نام MIFS-U ارائه شده است که تخمین بهتری از اطلاعات متقابل بین ویژگی‌های ورودی و برچسب کلاس به‌منظور افزایش کارایی روش MIFS ارائه داده است. این روش از تابع هدف زیر برای رتبه‌دهی ویژگی‌ها استفاده کرده است:

$$I(f_i; C) - \beta \sum_{f_s \in S} \frac{I(f_i; C)}{H(f_s)} I(f_s; f_i) \quad (12)$$

در روش [۲۳] که تعمیم یافته روش MIFS است، پارامتر β به‌صورت زیر حذف شده است:

$$I(f_i; C) - \frac{1}{|S|} \sum_{f_s \in S} I(f_s; f_i) \quad (13)$$

در مقاله [۳۷] نسخه تعمیم یافته MRMR با نام NMIFS ارائه شده است که در آن از اطلاعات متقابل نرمال شده به‌جای اطلاعات متقابل در محاسبه افزونگی ویژگی‌ها استفاده شده است. تابع هدف این روش به‌صورت زیر است:

$$I(X, Y | Z) = \sum_{x \in X} \sum_{y \in Y} \sum_{z \in Z} p(x, y, z) \log \frac{p(x, y | z)}{p(x | z) p(y | z)} \quad (7)$$

لم ۲. میان اطلاعات متقابل شرطی و اطلاعات متقابل رابطه زیر برقرار است:

$$I(X; Y | Z) = I(X; Y, Z) - I(X; Z) \quad (8)$$

اثبات: در مرجع [۵۱] ارائه شده است.

۲-۲- بهینه‌سازی چندهدفه

در مسائل بهینه‌سازی چندهدفه، هدف اصلی پیدا کردن بهترین جواب از میان همه جواب‌های امکان‌پذیر است. با توجه به تعداد توابع هدف، می‌توان آن‌ها را به دو دسته یک هدفه و چندهدفه تقسیم کرد. مسائل بهینه‌سازی چندهدفه، دو یا چند تابع هدف متضاد را به‌صورت هم‌زمان از طریق فرایندهای جستجو محاسبه می‌کند. یک مسئله بهینه‌سازی با چند تابع هدف را می‌توان به‌صورت زیر مدل کرد:

$$\min F(X) = [f_1(x), f_2(x), \dots, f_n(x)] \quad (9)$$

بطوریکه x به بردار متغیرهای تصمیم، $f_i(x)$ به تابع هدف i -ام به ازای x و متغیر M به تعداد توابع هدف اشاره می‌کند. از طرفی چون توابع با یکدیگر در تضاد هستند پیدا کردن جوابی که همه توابع را بهینه کند عملاً سخت است. در این شرایط، مجموعه‌ای از جواب‌های قابل قبول وجود دارد که راه‌حل‌های پارتو نامیده می‌شوند. به‌طور کلی، یک جواب غیر غالب یا بهینه پارتو نامیده می‌شود اگر هیچ جواب دیگری بر آن غالب نشود. از تعریف زیر به‌منظور جواب بهینه پارتو یا غالب استفاده می‌شود:

تعریف ۳: اگر جواب‌های یک مسئله دارای m تابع هدف باشند، آنگاه جواب x بر جواب y غالب (چیره) است اگر و تنها اگر شرایط زیر برقرار باشد:

$$\begin{aligned} \forall k = 1, \dots, m \quad f_k(x) &\leq f_k(y) \\ \exists k = 1, \dots, m \quad f_k(x) &< f_k(y) \end{aligned} \quad (10)$$

۲-۳- راهکارهای پیشین

تاکنون راهکارهای مختلفی برای انتخاب ویژگی ارائه شده است. این راهکارها به دو دسته راهکارهای پوششی و فیلتر دسته‌بندی می‌شوند. در راهکار پوششی از الگوریتم‌های یادگیری ماشین برای ارزیابی زیرمجموعه‌های انتخاب شده استفاده می‌شود. به همین دلیل روش‌های انتخاب ویژگی پوششی، دقت بالایی در انتخاب ویژگی‌ها با اهمیت دارند [۲۶، ۲۷]. اما به دلیل استفاده از الگوریتم یادگیری، پیچیدگی محاسباتی بالایی دارند و عملاً برای کاربردهای واقعی مانند دسته‌بندی متن که با مجموعه داده‌های بزرگی مواجه هستیم، مناسب نیستند. در مقابل در راهکار فیلتر، از روش‌های ارائه شده در نظریه اطلاعات برای ارزیابی زیرمجموعه‌های انتخابی استفاده می‌شود. لذا روش‌های انتخاب ویژگی از نوع فیلتر سرعت محاسباتی بالاتری نسبت به روش‌های پوششی دارند [۳۱-۲۸]. راهکارهای انتخاب ویژگی فیلتر خود به دو

متقابل شرطی ارائه شده است. شبه کد الگوریتم پیشنهادی در شکل ۱ ارائه شده است.

در ابتدا که هنوز هیچ ویژگی انتخاب نشده است، الگوریتم اطلاعات متقابل میان تک تک ویژگی‌ها و برجسب را محاسبه می‌کند و ویژگی که دارای بیشترین اطلاعات متقابل است در مجموعه S قرار می‌دهد (خطوط ۳ و ۴). به‌طور کلی می‌توان گفت، بعد از این مرحله الگوریتم پیشنهادی وارد دو فاز اصلی می‌شود. در فاز اول، میزان ارتباط و افزونگی به ازای همه ویژگی‌های کاندید (ویژگی‌هایی که هنوز انتخاب نشده‌اند) محاسبه شده و در فاز دوم، از میان این ویژگی‌ها، یک ویژگی با ارتباط بیشینه و افزونگی حداقل به‌منظور اضافه شدن در مجموعه هدف انتخاب می‌شود. این دو فاز تا زمانی که به تعداد دلخواه ویژگی انتخاب شود ادامه می‌یابد. در ادامه جزییات این فازها آورده شده است.

CJMIM: Conditional Joint Mutual Information Maximization	
Input:	F : Initial set of features, k : Size of selected set of features
Output :	S : The final set of selected features
1. Begin	
2.	$S = \{ \}$
3.	For each candidate feature f in F compute $I(f, C)$ End
4.	Find the feature f with maximum $I(f, C)$ and add to S
5.	Repeat until $ S =k$
6.	$F_i \leftarrow F \setminus S$
7.	For each feature f_i in F_i compute relevancy and redundancy based on equations (19) and (21)
8.	Objective_func(f_i) ← [relevancy(f_i), redundancy(f_i)]
9.	End for
10.	$[F_i] \leftarrow \text{NonDominateSort}(F_i)$
11.	$[F_i] \leftarrow \text{CrowdingDistance}(F_i)$
12.	Select $F_i(1)$ and to S
13.	End Repeat
14.	End Algorithm

شکل ۱: الگوریتم انتخاب ویژگی CJMIM

۳-۱- فاز اول: محاسبه ارتباط و افزونگی ویژگی‌های کاندید

در این فاز (خطوط ۷ تا ۹ الگوریتم)، به ازای همه ویژگی‌های کاندید، میزان ارتباط و افزونگی ویژگی‌ها با در نظر گرفتن ویژگی‌های انتخاب شده پیشین محاسبه می‌شود.

۳-۱-۱- محاسبه ارتباط یک ویژگی: میزان ارتباط یک ویژگی

کاندید f_i ، با برجسب کلاس C و با در نظر گرفتن سایر ویژگی‌های انتخاب شده قبلی در مجموعه $S = \{f_1, f_2, \dots, f_{|S|}\}$ با استفاده از اطلاعات متقابل شرطی $I(f_i, C|S)$ محاسبه می‌شود. طبق لم ۲، طریق زیر محاسبه می‌شود:

$$I(C; f_i | S) = I(f_i, S; C) - I(C; S) \quad (15)$$

با در نظر گرفتن $S' = \{S, f_i\}$ مقدار $I(S'; C)$ از رابطه زیر به دست می‌آید:

$$I(f_i; C) - \frac{1}{|S|} \sum_{f_j \in S} \frac{I(f_j; f_i)}{\min\{H(f_j), H(f_i)\}} \quad (14)$$

در [۳۸] یک روش انتخاب ویژگی چندهدفه به نام MIFS_ND ارائه شده که با استفاده از بیشینه‌سازی اطلاعات متقابل بین کلاس و ویژگی‌ها و کمینه‌سازی متوسط اطلاعات متقابل میان ویژگی‌ها به انتخاب ویژگی‌های مهم پرداخته است. در [۴۰] یک روش انتخاب ویژگی با نام JMI ارائه شده است که بر مبنای آن ویژگی‌هایی که منجر به افزایش مجموع اطلاعات متقابل مشترک با ویژگی‌های موجود شوند، انتخاب می‌شوند. این روش از دقت دسته‌بندی و پایداری بالایی برخوردار است. علاوه بر این در [۴۶] یک روش انتخاب ویژگی فیلتر به نام ECFS ارائه شده است که با استفاده از بردارهای ویژه مرکزی^۷ مسئله انتخاب ویژگی را به گراف وابستگی تبدیل می‌کند. در این روش، ویژگی‌ها به گره‌ها نگاشت می‌شوند و سپس با استفاده از بعضی شاخص‌های مرکزی، مانند بردار ویژه اهمیت ویژگی‌ها تخمین زده می‌شود. اساس این روش این است که میزان اهمیت هر ویژگی بر اساس اهمیت همسایه‌های آن محاسبه می‌شود. در [۴۷] نیز یک روش انتخاب ویژگی فیلتر به نام infFS پیشنهاد شده است. این روش بر اساس مفهوم مرکزیت و انتگرال مسیر روی گراف همسایگی، ویژگی‌ها را امتیازدهی می‌کند. از خصوصیات جالب این روش می‌توان ارزیابی تمام حالات احتمالی ممکن برای ویژگی‌ها دانست. علاوه بر این از آنجا که مسئله انتخاب ویژگی را به محاسبه سری هندسی ماتریس همسایگی^۸ تبدیل می‌کند از لحاظ زمانی نیز بسیار کارآمد است. علاوه بر نظریه اطلاعات، از نظریه مجموعه‌های سخت [۳۴، ۵۰] و تحلیل پوششی داده [۳۵] در رتبه‌بندی ویژگی‌ها نیز استفاده شده است. یکی از مشکلات استفاده از نظریه مجموعه‌های سخت و تحلیل پوششی داده‌ها، پیچیدگی محاسباتی بالای آن‌ها است. روش‌هایی که تاکنون مورد بررسی قرار گرفتند برای انتخاب ویژگی در داده‌های تک برجسبی بودند. روش‌های انتخاب ویژگی برای داده‌های با چند برجسب نیز ارائه شده است. به‌عنوان مثال در [۴۵] یک روش انتخاب ویژگی چند برجسبی بر اساس نظریه اطلاعات ارائه شده است. در این روش یک تابع امتیازدهی جدید برای تعیین میزان اهمیت هرکدام از ویژگی‌ها نسبت به چندین برجسب تعریف شده است. همچنین در [۴۹] یک روش انتخاب ویژگی چند برجسبی دیگر ارائه شده که در آن از ترکیب یک روش فیلتر و پوششی استفاده شده است.

۳- روش پیشنهادی

هدف اصلی این مقاله ارائه روشی برای انتخاب زیرمجموعه‌ای از ویژگی‌هاست که دارای بیشترین ارتباط و کمترین افزونگی باشند. همان‌طور که قبلاً نیز اشاره شد، ارتباط یک ویژگی به میزان وابستگی آن با برجسب کلاس و افزونگی یک ویژگی به میزان وابستگی میان ویژگی‌ها مربوطه با سایر ویژگی‌های انتخاب شده اشاره دارد. به همین منظور در این مقاله روشی با عنوان CJMIM با دو تابع هدف یکی برای بیشینه کردن مقدار ارتباط و یکی برای کمینه کردن مقدار افزونگی با استفاده از اطلاعات

۳-۱-۲- محاسبه میزان افزونگی یک ویژگی: از طرف دیگر، میزان افزونگی یک ویژگی با استفاده از اطلاعات متقابل شرطی میان ویژگی کاندید و سایر ویژگی‌های قبلاً انتخاب شده دیگر و با در نظر گرفتن برچسب کلاس C با استفاده از $I(f_i, S|C)$ محاسبه می‌شود.

تعریف ۴. افزونگی ویژگی کاندید $f_i \in F \setminus S$ نسبت به سایر ویژگی‌های کاندید $f_j \in F \setminus S$ بیشتر است اگر $I(f_i, S|C) > I(f_j, S|C)$ بطوریکه $I(f_i, S|C)$ نشان‌دهنده اطلاعات متقابل شرطی میان ویژگی f_i و مجموعه ویژگی‌های انتخاب شده S با در نظر گرفتن برچسب کلاس C است. از آنجا که محاسبه این مقدار طبق رابطه (۱۶) زمان‌بر است پس مقدار این عبارت نیز مانند زیر تخمین زده می‌شود.

لم ۴. با در نظر گرفتن مجموعه $F = \{f_1, f_2, \dots, f_N\}$ ، $S \subseteq F$ و $f_i \in F \setminus S$ به ترتیب به‌عنوان مجموعه ویژگی‌های اولیه، مجموعه ویژگی‌های انتخاب شده و ویژگی کاندید، مقدار افزونگی ویژگی کاندید f_i با در نظر گرفتن برچسب کلاس C و S عبارت است از:

$$I_{f_i \in F \setminus S}(f_i, S|C) = \text{mean}_{f_j \in S} I(f_i, f_j|C) \quad (21)$$

اثبات: طبق لم شماره ۲ مقدار $I(f_i, S|C)$ عبارت است از

$$I(f_i, S|C) = I(S; f_i|C) = I(S; f_i, C) - I(S; C) \quad (22)$$

حال با جایگذاری روابط (۱۷) و (۱۸) در (۲۱) رابطه زیر به دست می‌آید:

$$\begin{aligned} I(f_i, S|C) &\equiv \frac{1}{|S|} \sum_{f_j \in S} I(f_i, f_j, C) - \frac{1}{|S|} \sum_{f_j \in S} I(f_j, C) \\ &\equiv \frac{1}{|S|} \sum_{f_j \in S} I(f_i, f_j|C) \equiv \text{mean}_{f_j} (I(f_i, f_j|C)) \end{aligned} \quad (23)$$

۳-۲- فاز دوم، انتخاب ویژگی: حال که به ازای هر کدام از ویژگی‌های کاندید دو مقدار ارتباط و افزونگی طبق روابط (۱۹) و (۲۱) به دست آمده است نوبت به انتخاب ویژگی با ارتباط بالا و افزونگی پایین می‌رسد. از آنجا که دو تابع هدف در تضاد با یکدیگر هستند پس از اصل بهینگی پارتو به‌منظور انتخاب ویژگی استفاده شده است. این فاز شامل سه مرحله است که در ادامه هر کدام از مراحل با جزییات ارائه شده است (خطوط ۱۰ تا ۱۲ الگوریتم).

۳-۲-۱- مرتب‌سازی غالب [۴۸]: در این مرحله (خط ۱۰) ویژگی‌ها براساس اصل بهینگی پارتو مرتب می‌شوند. طبق این الگوریتم رتبه یک متعلق به جواب‌هایی است که توسط هیچ جواب دیگری چیره نمی‌

$$\begin{aligned} I(S'; C) &= \sum \sum p(S', C) \log \frac{p(S', C)}{P(S')P(C)} \\ &= \sum \sum p(f_1, f_2, \dots, f_{|S'|}, C)^* \\ &\quad \log \frac{p(f_1, f_2, \dots, f_{|S'|})}{p(f_1, f_2, \dots, f_{|S'|})P(C)} \end{aligned} \quad (16)$$

همان‌طور که مشاهده می‌شود پیچیدگی محاسباتی رابطه بالا با افزایش تعداد ویژگی‌ها و نمونه‌های آموزشی افزایش می‌یابد. به همین دلیل محاسبه دقیق اطلاعات متقابل شرطی با استفاده از فرمول بالا زمان‌بر است. در [۲۳] یک تابع به‌منظور تخمین اطلاعات متقابل بنام MRMR ارائه شده است که اطلاعات متقابل میان یک مجموعه از ویژگی‌های S و ویژگی C را با استفاده از میانگین اطلاعات متقابل میان تک تک ویژگی‌های مجموعه S و ویژگی C به دست آورده است. بر این اساس در این مقاله یک الگوریتم انتخاب ویژگی جدید ارائه می‌شود که با استفاده از [۲۳] مقادیر $I(S, C)$ و $I(f_i, S; C)$ در رابطه (۱۵) را همانند زیر محاسبه می‌کند:

$$I(f_i, S; C) = \frac{1}{|S|} \sum_{f_j \in S} I(f_i, f_j; C) \quad (17)$$

$$I(S; C) = \frac{1}{|S|} \sum_{f_j \in S} I(f_j; C) \quad (18)$$

لم ۳. اگر مجموعه $F = \{f_1, f_2, \dots, f_N\}$ و $S \subseteq F$ به ترتیب به مجموعه ویژگی‌های اولیه و مجموعه ویژگی‌های انتخاب شده اشاره کند و $f_i \in F \setminus S$ ، آنگاه می‌توان گفت معیار ارتباط ویژگی کاندید f_i و برچسب کلاس C با وجود ویژگی‌های قبلاً انتخاب شده در S به‌صورت زیر محاسبه می‌شود:

$$I_{f_i \in F \setminus S}(f_i, C|S) = \text{mean}_{f_j \in S} I(f_i, C|f_j) \quad (19)$$

اثبات: با جایگذاری روابط (۱۷) و (۱۸) در رابطه شماره (۱۵)، رابطه زیر به دست می‌آید:

$$\begin{aligned} I(f_i, C|S) &\equiv \frac{1}{|S|} \sum_{f_j \in S} I(f_i, f_j, C) - \frac{1}{|S|} \sum_{f_j \in S} I(f_j, C) \\ &\equiv \frac{1}{|S|} \sum_{f_j \in S} (I(f_i, f_j, C) - I(f_j, C)) \\ &\equiv \frac{1}{|S|} \sum_{f_j \in S} I(f_i, C|f_j) \\ &\equiv \text{mean}_{f_j} (I(f_i, C|f_j)) \end{aligned} \quad (20)$$

صورت می‌گیرد. تمام آزمایش‌های انجام شده روی سیستمی با حافظه چهار گیگابایت و پردازشگر از نوع AMD A8 انجام شده است.

در آزمایش‌های انجام شده بعضی روش‌ها مانند MOEA/D با اعمال بر روی داده‌های حجیم کارایی خود را از دست می‌دهند. به‌عنوان مثال با اعمال این الگوریتم به داده آموزشی ColonTumor با ۲۰۰۰ ویژگی معمولاً نزدیک به ۷۰۰ ویژگی انتخاب می‌شود. از آنجا که این تعداد ویژگی معمولاً زیاد است و کاربردی نیست و همچنین انتخاب این تعداد ویژگی در الگوریتم‌های فیلتر معرفی شده در این مقاله که براساس نظریه اطلاعات متقابل ویژگی‌ها را انتخاب می‌کنند مستلزم صرف زمان زیادی است پس از اعمال این الگوریتم روی داده‌های حجیم صرف‌نظر می‌شود. در همین راستا نتایج به‌دست‌آمده در قالب دقت دسته‌بندی در دو قسمت ارائه شده است. در قسمت اول، الگوریتم‌های انتخاب ویژگی روی داده‌های آموزشی با تعداد ویژگی کم مانند Pima-Flags, Hepatit, Vehicle, Diabetes, Heart-statlog, Credit-a و Credit-g بر روی همه الگوریتم‌های معرفی شده در مقاله باهم مقایسه شده‌اند. در بخش دوم نیز به‌جز الگوریتم MOEA/D سایر الگوریتم‌ها روی داده‌های آموزشی با تعداد ویژگی زیاد و یا با تعداد نمونه‌های آموزشی زیاد مانند Hypothyroid, ColonTumor, SRBCT, CNS و Lymphoma باهم مقایسه شده‌اند.

جدول ۱: مشخصات داده‌های آموزشی

نام داده‌ها	تعداد ویژگی‌ها	تعداد نمونه‌ها
Hepatitis	۱۹	۱۵۵
Flags	۲۹	۱۹۴
Pima-diabetes	۸	۷۶۸
Vehicle	۱۸	۸۴۶
Heart-statlog	۱۳	۲۷۰
Creat-a	۱۵	۶۹۰
Credit-g	۲۰	۱۰۰۰
Hypothyroid	۲۹	۳۷۷۹
colonTumor	۲۰۰۰	۶۲
SRBCT	۲۳۰۸	۸۳
CNS	۷۱۲۹	۶۰
Lymphoma	۴۰۲۶	۶۶

۴-۱- نتایج به‌دست‌آمده روی داده‌های آموزشی در مقیاس متوسط

در این بخش، نتایج به‌دست‌آمده به ازای روش پیشنهادی و سایر روش‌های انتخاب ویژگی از جمله ECFS, MIFS, MRMR, MIFS-ND, ECFS, MOEA/D و InfFS روی داده‌های آموزشی متوسط که در جدول شماره ۱ معرفی شده‌اند در شکل‌های شماره ۲ تا ۸ به ازای هر یک از

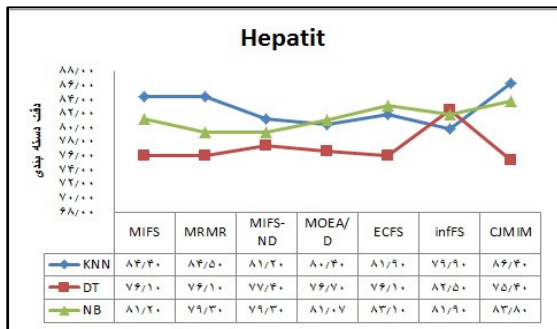
شوند (به عبارتی دیگر بهینه پارتو هستند) و رتبه دو به آن دسته از ویژگی‌هایی تعلق می‌گیرد که با حذف ویژگی‌های رتبه اول از مجموعه ویژگی‌ها، بر سایر ویژگی‌های باقیمانده غلبه کنند. بر همین اساس رتبه ۳ و ۴ والی آخر به ازای همه ویژگی‌های دیگر نیز محاسبه شده و در نهایت ویژگی‌هایی که دارای رتبه یکسانی هستند در یک مجموعه قرار داده می‌شوند. در این نوع رتبه‌بندی، ویژگی‌هایی که رتبه کمتری دارند نسبت به سایر ویژگی‌ها با رتبه بالاتر بهتر هستند.

۳-۲-۲- محاسبه فاصله جمعیتی [۴۸]: در این مرحله (خط ۱۱)، به ازای هر مجموعه از ویژگی‌ها که دارای رتبه پارتو یکسانی هستند، معیار فاصله جمعیتی محاسبه می‌شود. فاصله جمعیتی یک جواب، تخمینی از چگالی جواب‌های نزدیک به آن است. این معیار را با استفاده از میانگین فاصله دو جواب که در همسایگی جواب اولیه هستند می‌توان محاسبه کرد.

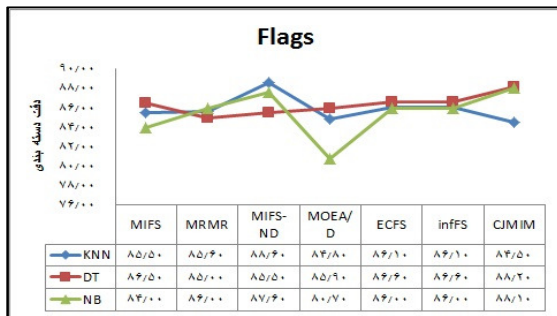
۳-۲-۳- انتخاب ویژگی: در این مرحله در ابتدا ویژگی‌ها براساس رتبه به‌صورت صعودی و سپس بر اساس فاصله جمعیتی به‌صورت نزولی مرتب می‌شوند. به این صورت ویژگی که در ابتدای لیست قرار می‌گیرد ویژگی انتخابی خواهد بود که دارای رتبه پایین‌تر و فاصله جمعیتی بالاتری است.

۴- نتایج آزمایشگاهی

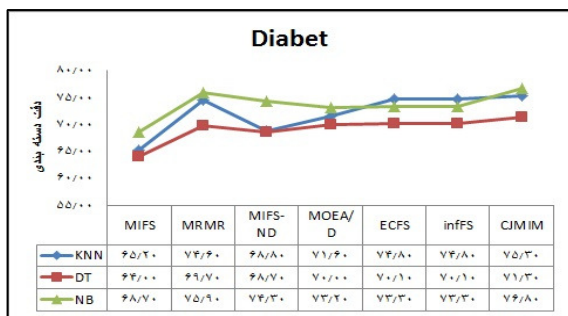
در این مقاله از چندین مجموعه داده، با مشخصات مختلف، برای ارزیابی روش پیشنهادی و مقایسه آن با سایر روش‌های انتخاب ویژگی استفاده شده است. مشخصات کلی این مجموعه‌های داده‌ای در جدول شماره ۱ ارائه شده است. برای بررسی کارایی روش پیشنهادی (CJMIM) از دسته‌بندی درخت تصمیم (DT)، k نزدیک‌ترین همسایه (KNN) و بیز ساده (NB) به‌منظور مقایسه با دیگر روش‌های انتخاب ویژگی از جمله ECFS, MRMR, MIFS, MIFS-ND, MOEA/D, infFS و ECFS استفاده شده است. از میان این روش‌ها، MOEA/D و MIFS-ND روش‌های انتخاب ویژگی چندهدفه هستند که به ترتیب جزء دسته روش‌های فیلتر و پوششی به‌حساب می‌آیند. از آنجا که تعداد ویژگی‌های انتخابی توسط الگوریتم MOEA/D متغیر است پس در ابتدا این روش را ۵ بار روی هر کدام از داده‌ها تست کرده و سپس سایر روش‌های انتخاب ویژگی را به ازای متوسط تعداد ویژگی‌های انتخاب شده توسط روش MOEA/D مقایسه می‌کنیم. در تمامی آزمایش‌ها، مجموعه‌های داده‌ای به‌صورت تصادفی به داده‌های آموزشی (۲/۳) از کل مجموعه داده‌ها و داده‌های آزمایشی تقسیم شدند. با توجه به تقسیم تصادفی داده‌ها، میانگین دقت دسته‌بندی بر روی داده‌های آزمایشی در اجراهای مختلف گزارش شده است. تقسیم داده‌ها در هر تکرار مستقل از تکرارهای دیگر



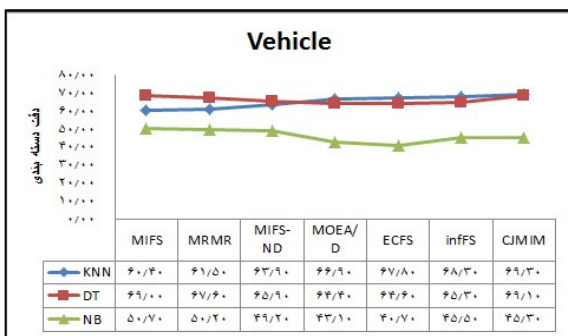
شکل ۲: دقت دسته‌بندی KNN، DT و NB به ازای $k=۸$



شکل ۳: دقت دسته‌بندی KNN، DT و NB به ازای $k=۹$



شکل ۴: دقت دسته‌بندی KNN، DT و NB به ازای $k=۳$



شکل ۵: دقت دسته‌بندی KNN، DT و NB به ازای $k=۱۰$

مجموعه‌های داده‌ای آورده شده است. محور افقی و عمودی در این شکل‌ها به ترتیب نشان‌دهنده یک الگوریتم انتخاب ویژگی و دقت دسته‌بندی است. همچنین در هرکدام از این شکل‌ها سه نمودار به رنگ‌های آبی، قرمز و سبز آورده شده که نشان‌دهنده دقت به‌دست‌آمده به ازای دسته‌بندی‌های KNN، DT و NB است. طبق نتایج به‌دست‌آمده در اکثر این شکل‌ها و به ازای بیشتر دسته‌بندی‌ها، سطح نمودار به ازای الگوریتم پیشنهادی که با نام اختصاری CJMIM آورده شده است، بیشتر است. به‌عنوان مثال به ازای شکل ۲ که مربوط به داده Hepatit است، الگوریتم پیشنهادی به ازای دسته‌بندی‌های KNN و NB با انتخاب ۸ ویژگی دقت بیشتری به دست آورده است. علاوه بر این نتایج به‌دست‌آمده به ازای داده Flags در شکل ۳ نشان می‌دهد که دقت دسته‌بندی‌های DT و NB در الگوریتم پیشنهادی با انتخاب ۹ ویژگی از سایر الگوریتم‌ها بالاتر است. همچنین در شکل ۴ که مربوط به داده آموزشی Diabetes است، می‌توان مشاهده کرد با انتخاب ۳ ویژگی، دقت دسته‌بندی KNN، DT و NB به‌دست‌آمده به ازای الگوریتم پیشنهادی از سایر روش‌های انتخاب ویژگی بالاتر است. به‌عنوان مثال دقت KNN به ازای الگوریتم پیشنهادی، MIFS، MRMR، MIFS-ND، ECFS، infFS و MOEA/D به ترتیب عبارت است از $۷۵/۳$ ، $۶۵/۲$ ، $۷۴/۶$ ، $۶۸/۸$ ، $۷۴/۸$ و $۷۱/۶$ است. در شکل‌های ۵، ۶، ۷ و ۸ به ترتیب نتایج به‌دست‌آمده به ازای داده‌های Vehicle، Heart-statlog، Credit-a و Credit-g نشان داده شده است. نتایج به ازای این مجموعه‌های داده‌ای نیز مشابه همان است که به ازای داده‌های قبلی به دست آمد. به‌عنوان مثال، نتایج به‌دست‌آمده به ازای Vehicle (شکل ۵) نشان می‌دهد که دقت KNN به ازای الگوریتم پیشنهادی از سایر الگوریتم‌ها بالاتر و به ازای دسته‌بندی DT دومین دقت را دارد. نتایج مربوط به داده Heart-statlog (شکل ۶) نشان می‌دهد که نتایج به‌دست‌آمده به ازای دسته‌بندی‌های KNN و DT مربوط به الگوریتم پیشنهادی از سایر الگوریتم‌ها بالاتر و دقت به‌دست‌آمده به ازای دسته‌بندی NB از الگوریتم‌های MOEA/D، ECFS و infFS بالاتر است. شکل ۷ نشان می‌دهد که دقت KNN مربوط به الگوریتم پیشنهادی با انتخاب ۶ ویژگی از سایر الگوریتم‌ها بالاتر است. در آخر در شکل ۸، مشاهده می‌شود که به ازای داده Credit-g و با انتخاب ۶ ویژگی، دقت دسته‌بندی‌های KNN، DT و NB در الگوریتم پیشنهادی بالاتر از سایر روش‌ها است.

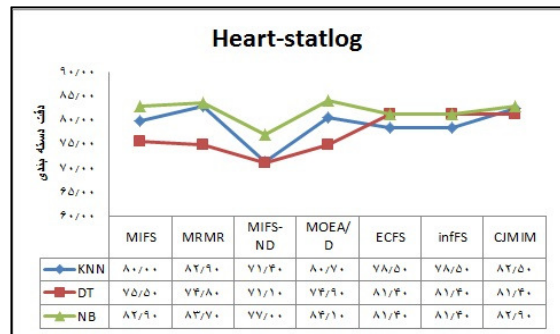
است به‌عنوان مثال به ازای $k=10$ ، دقت DT مربوط به الگوریتم پیشنهادی، MIFS، MRMR، MIFS-ND، ECFS و infFS به ترتیب عبارت است از $99/0.8$ ، $97/2$ ، $96/1$ ، $99/0.2$ و $92/2$ که دقت الگوریتم پیشنهادی از میان همه الگوریتم‌های دیگر بالاتر است. نتایج مربوط به داده ColonTumor در جدول ۳ نشان داده شده است. نتایج نشان می‌دهد که الگوریتم پیشنهادی به ازای دسته‌بند KNN، از چهار مورد در سه مورد بیشترین دقت را به دست آورده است. همچنین به ازای دسته بند DT در یک مورد بیشترین دقت و در بقیه موارد معمولاً یا دومین بهترین و یا سومین بهترین را از میان ۶ الگوریتم انتخاب ویژگی به دست آورده است. نتایج مربوط به داده SRBCT در جدول ۴ نشان می‌دهد که الگوریتم پیشنهادی به ازای دسته‌بندهای KNN و DT در ۳ مورد از چهار مورد بیشترین دقت را به دست آورده است. نتایج به‌دست‌آمده به ازای داده CNS در جدول شماره ۵ نشان می‌دهد که الگوریتم پیشنهادی در دو مورد و به ازای دسته‌بند DT در سه مورد از چهار مورد بیشترین دقت را به دست آورده است. نتایج به‌دست‌آمده به ازای داده Lymphoma در جدول ۶ نشان می‌دهد که الگوریتم پیشنهادی به ازای دسته بند KNN در سه مورد بهترین دقت و در یک مورد باقیمانده دومین بهترین دقت را داشته است. از طرف دیگر به ازای دسته‌بند DT معمولاً چهارمین یا سومین دقت را به دست آورده است. نتایج مربوط به داده Splice در جدول ۷ نیز نشان می‌دهد که الگوریتم پیشنهادی معمولاً به ازای همه k ها و دسته‌بندها معمولاً بیشترین دقت را به دست آورده است.

جدول ۲: متوسط دقت دسته‌بندهای KNN و DT به ازای Hypothyroid

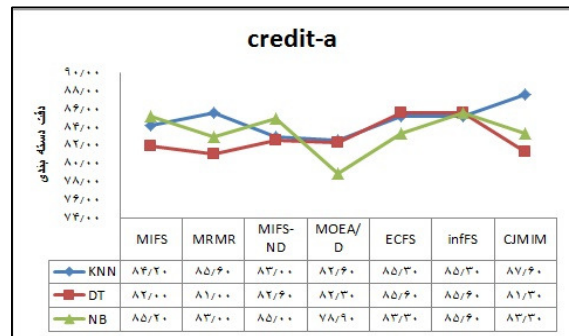
	k	MIFS	MRMR	MIFS-ND	CJMIM	ECFS	infFS
KNN	۶	۹۶/۰۷	۹۵/۹۱	۹۵/۸۱	۹۴/۴۳	۹۱/۹۶	۹۱/۹۶
	۸	۹۶/۲۸	۹۵/۸۳	۹۵/۷۸	۹۴/۳۷	۹۱/۹۶	۹۱/۹۶
	۱۰	۹۶/۱۵	۹۵/۲	۹۴/۵۱	۹۴/۴۰	۹۲/۲۳	۹۲/۲۳
	۱۲	۹۶/۲۳	۹۵/۰۲	۹۴/۶۱	۹۴	۹۲/۲۵	۹۲/۲۵
DT	۶	۹۶/۷۳	۹۵/۹۷	۹۷/۷۴	۹۷/۱۱	۹۲/۲۸	۹۲/۲۸
	۸	۹۷	۹۵/۹۹	۹۷/۷۹	۹۹/۱	۹۲/۲۵	۹۲/۲۵
	۱۰	۹۷/۲۶	۹۶/۱۵	۹۹/۰۲	۹۹/۰۸	۹۲/۲۸	۹۲/۲۸
	۱۲	۹۷/۰۵	۹۶/۱۲	۹۸/۷	۹۸/۷۲	۹۲/۱۷	۹۲/۱۷

جدول ۳: متوسط دقت دسته‌بندهای KNN و DT به ازای ColonTumor

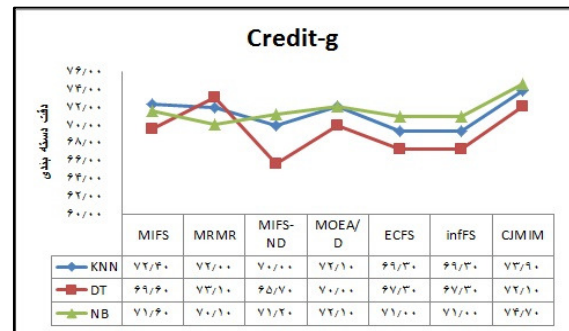
	k	MIFS	MRMR	MIFS-ND	CJMIM	ECFS	infFS
KNN	۶	۸۷/۰۶	۷۸/۰۴	۸۶/۹۸	۸۷/۱۴	۷۲/۶۱	۷۴/۲۰
	۸	۸۷/۰۲	۷۷/۴۶	۸۸/۶۵	۸۸/۷۳	۷۹/۱۲	۷۵/۸۷
	۱۰	۸۷/۲۲	۷۲/۵۳	۸۵/۵۳	۸۵/۵۵	۸۳/۸۸	۷۹/۰۴
	۱۲	۸۰/۰۶	۷۴/۳۶	۸۳/۷۰	۸۵/۴۰	۷۵/۸۴	۷۷/۴۱
DT	۶	۸۶/۹۰	۷۹/۰۴	۸۶/۹۰	۸۳/۸۸	۶۱/۴۲	۶۳/۱۵
	۸	۸۳/۹۶	۷۲/۳۸	۸۳/۹۶	۸۳/۳۸	۷۲/۶۹	۷۵/۷۱
	۱۰	۸۷/۲۲	۷۰/۹۵	۸۸/۸۰	۸۰/۷۹	۶۶/۰۳	۶۵/۹۵
	۱۲	۷۴/۰۵	۶۴/۷۱	۷۴/۰۵	۷۹/۱۸	۶۹/۴۴	۷۴/۱۳



شکل ۶: دقت دسته‌بندی KNN، DT و NB به ازای $k=6$



شکل ۷: دقت دسته‌بندی KNN، DT و NB به ازای $k=6$



شکل ۸: دقت دسته‌بندی KNN، DT و NB به ازای $k=8$

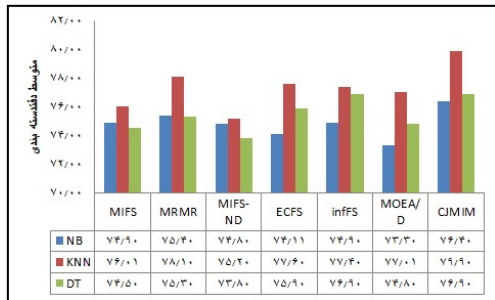
۴-۲- نتایج به‌دست‌آمده بر روی داده آموزشی در مقیاس بزرگ

به‌منظور مقایسه بیشتر روش پیشنهادی با سایر روش‌های انتخاب ویژگی، آزمایش‌ها بر روی ۶ داده آموزشی حجیم تکرار شده است. در ادامه نتایج به‌دست‌آمده به ازای سایر الگوریتم‌ها و الگوریتم پیشنهادی به ازای انتخاب ۶، ۸، ۱۰ و ۱۲ ویژگی و با استفاده از دسته‌بندهای KNN و DT در جدول‌های ۲ تا ۷ نشان داده شده است.

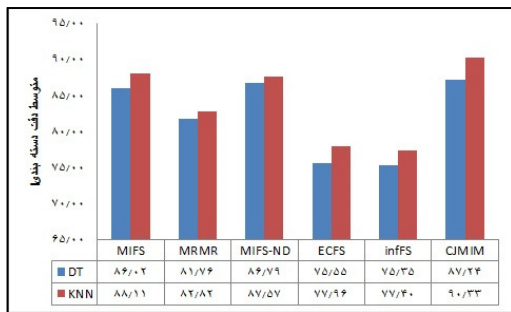
هرکدام از جدول‌ها به ازای یک مجموعه داده‌ای می‌باشند که ستون‌ها نشان‌دهنده یک الگوریتم و سطرها به ازای تعداد ویژگی‌های مختلف و دسته‌بندهای KNN و DT است. در هر سطر مقادیر بیشینه به‌صورت زیرخط دار نشان داده شده است. نتایج به‌دست‌آمده به ازای این مجموعه داده‌ها نیز تقریباً مشابه نتایجی است که به ازای مجموعه داده‌های معمولی به دست آمد. همان‌طور که در جدول ۲ (این جدول مربوط به داده Hypothyroid می‌باشد) مشاهده می‌شود دقت الگوریتم پیشنهادی به ازای دسته‌بند DT و همه k ها از سایر الگوریتم‌ها بیشتر

۵- بحث

نتایج به‌دست‌آمده در بخش‌های قبلی با استفاده از دسته‌بندهای KNN، DT، NB و به ازای تعداد ویژگی‌های مختلف مورد مقایسه قرار گرفت. نتایج ارائه شده نشان داد که دقت الگوریتم پیشنهادی با استفاده از دسته‌بندهای KNN، DT و NB در بیشتر موارد و به ازای اکثر مجموعه داده‌ها بیشتر از سایر الگوریتم‌ها است. علاوه بر این میانگین و رتبه الگوریتم پیشنهادی روی داده‌های مختلف در مقایسه با سایر الگوریتم‌های انتخاب ویژگی با استفاده از آزمون فریدمن در شکل‌های ۹ تا ۱۲ نشان داده شده است. هرچقدر میانگین دقت دسته‌بندی یک الگوریتم بیشتر و رتبه یک الگوریتم به ازای همه مجموعه‌های داده‌ای کمتر باشد مطلوبیت الگوریتم بیشتر است. همان‌طور که در شکل ۹ و ۱۰ مشاهده می‌شود متوسط دقت دسته‌بندی الگوریتم پیشنهادی به ازای همه دسته‌بندها بیشتر از سایر الگوریتم‌ها است. همچنین رتبه الگوریتم‌ها با استفاده از دسته‌بندهای مختلف در شکل‌های ۱۱ و ۱۲ نشان داده شده است. شکل ۱۱ مربوط به رتبه الگوریتم‌های انتخاب ویژگی روی داده‌های با اندازه متوسط است.



شکل ۹: میانگین دقت دسته‌بندها به ازای الگوریتم‌های انتخاب ویژگی مختلف روی داده‌های آموزشی با سایز متوسط



شکل ۱۰: میانگین دقت دسته‌بندها به ازای الگوریتم‌های انتخاب ویژگی مختلف روی داده‌های با سایز بزرگ

جدول ۴: متوسط دقت دسته‌بندهای KNN و DT به ازای SRBCT

	k	MIFS	MRMR	MIFS-ND	CJMIM	ECFS	infFS
KNN	۶	۸۶/۸۱	۷۱/۲۰	۸۸	۹۲/۷۶	۶۷/۵۴	۶۷/۵۴
	۸	۸۵/۵۳	۶۵/۰۷	۸۵/۵۳	۸۶/۷۷	۶۹/۸۴	۶۹/۸۴
	۱۰	۸۹/۱۹	۷۳/۵۸	۸۹/۱۰	۸۵/۵۳	۶۵/۱۶	۶۵/۱۶
	۱۲	۸۴/۲۵	۷۳/۵۶	۸۶/۶۴	۸۶/۶۴	۷۱/۱۵	۷۱/۱۵
DT	۶	۷۷/۲۰	۶۷/۵۹	۷۷/۲۰	۸۲/۰۵	۶۷/۳۲	۶۷/۳۲
	۸	۸۳/۱۵	۶۲/۶۵	۸۳/۱۵	۸۲/۱۱	۶۸/۶۹	۶۸/۶۹
	۱۰	۸۴/۴۳	۶۷/۶۳	۸۴/۴۳	۸۱/۹۶	۷۴/۶۹	۷۴/۶۹
	۱۲	۸۰/۲۶	۷۲/۵۱	۸۰/۶۹	۸۳/۷۹	۶۹/۱۵	۶۹/۱۵

جدول ۵: متوسط دقت دسته‌بندهای KNN و DT به ازای داده CNS

	k	MIFS	MRMR	MIFS-ND	CJMIM	ECFS	infFS
KNN	۶	۷۱/۷۴	۸۶/۶۵	۷۱/۷۴	۷۳/۳۲	۷۰/۰۵	۷۱/۷۸
	۸	۶۸/۳۵	۶۳/۳۵	۷۰/۰۹	۷۸/۳۶	۶۸/۳۵	۷۳/۳۶
	۱۰	۸۷/۹۱	۸۹/۴۱	۸۹/۴۱	۸۷/۹۵	۸۷/۹۵	۸۳/۳۹
	۱۲	۸۶/۷۴	۶۳/۳۷	۸۱/۷۵	۷۶/۷۴	۶۵/۰۵	۵۸/۷۹
DT	۶	۶۸/۹۵	۶۱/۷۲	۶۸/۳۴	۶۸/۳۴	۶۸/۳۴	۶۶/۷۴
	۸	۵۳/۳۶	۵۳/۳۶	۵۳/۳۶	۶۳/۴۷	۶۰/۰۵	۶۱/۷۸
	۱۰	۸۷/۹۴	۸۹/۴۲	۸۹/۴۱	۸۷/۹۰	۸۷/۹۰	۸۳/۳۱
	۱۲	۶۵/۰۴	۶۱/۷۹	۶۵/۰۸	۷۶/۷۴	۵۶/۷۹	۵۱/۷۹

جدول ۶: متوسط دقت دسته‌بندهای KNN و DT به ازای Lymphoma

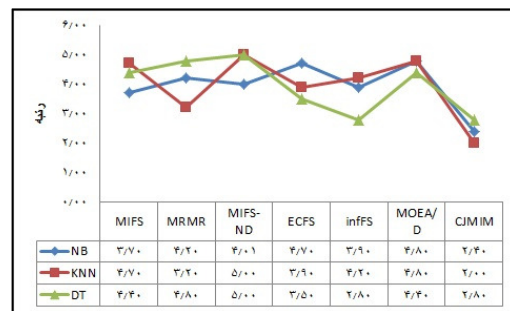
	k	MIFS	MRMR	MIFS-ND	CJMIM	ECFS	infFS
KNN	۶	۹۰/۰۲	۸۴/۸۷	۹۸/۵۶	۹۷/۰۴	۸۱/۸۰	۸۱/۸۰
	۸	۹۵/۵۴	۸۶/۴۱	۹۳/۹۵	۹۸/۵۰	۷۸/۸۴	۷۸/۸۱
	۱۰	۹۲/۴۱	۸۷/۹۶	۹۷/۰۴	۹۷/۰۵	۸۹/۴۵	۸۴/۷۸
	۱۲	۸۷/۰۴	۹۳/۹۹	۹۳/۹۹	۱۰۰	۹۲/۴۱	۹۲/۴۵
DT	۶	۸۶/۴۱	۸۶/۴۵	۸۹/۴۱	۸۴/۸۷	۷۴/۲۲	۷۴/۲۲
	۸	۸۶/۴۱	۸۷/۷۷	۸۷/۸۴	۸۴/۷۱	۸۰/۳۵	۸۰/۳۳
	۱۰	۸۷/۹۶	۸۹/۴	۸۹/۴	۸۷/۹۴	۸۷/۹۴	۸۳/۳۵
	۱۲	۸۳/۳۵	۹۰/۹۶	۸۷/۹۴	۸۷/۹۴	۹۰/۰۹	۹۰/۰۹

جدول ۷: متوسط دقت دسته‌بندهای KNN و DT به ازای Splice

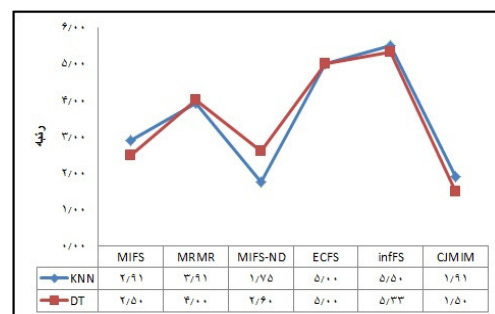
	k	MIFS	MRMR	MIFS-ND	CJMIM	ECFS	infFS
KNN	۶	۹۲/۳	۹۶/۲۶	۸۳/۷۹	۹۶/۶۱	۷۱/۷۵	۷۱/۷۵
	۸	۹۲/۳	۹۶/۲۶	۸۳/۷۹	۹۶/۶۱	۷۱/۷۵	۷۱/۷۵
	۱۰	۹۰/۲	۹۵/۳۹	۸۲/۳۵	۹۵/۶۷	۷۰/۱۸	۷۰/۱۸
	۱۲	۸۸/۱	۹۳/۹۴	۸۰/۴۷	۹۴/۲۳	۷۰/۰۰	۷۰/۰۶
DT	۶	۹۹/۹	۹۹/۹۶	۹۹/۹۶	۹۹/۹۶	۷۲/۳۸	۷۲/۳۸
	۸	۹۹/۹	۹۹/۹۶	۹۹/۹۶	۹۹/۹۶	۷۱/۶۳	۷۲/۶۳
	۱۰	۹۹/۸	۹۹/۸۴	۹۹/۸۴	۹۹/۸۴	۷۱/۹۴	۷۱/۸۱
	۱۲	۹۹/۹	۹۹/۹۶	۹۹	۹۹/۹۶	۷۳/۱۹	۷۳/۱۹

مراجع

- [1] H. Liu and H. Motoda, *Computational Methods of Feature Selection (Chapman & Hall/CRC Data Mining and Knowledge Discovery Series)*, Chapman & Hall/CRC, 2007
- [2] I. T. Jolliffe, *Principal Component Analysis*, Springer-Verlag New York, 2002.
- [3] S. Mika, G. Ratsch, J. Weston, B. Scholkopf, K.R. Mullers, "Fisher discriminant analysis with kernels," *Neural Networks for Signal Processing IX: Proceedings of the 1999 IEEE Signal Processing Society Workshop (Cat. No.98TH8468)*, Instituted of Electrical and Electronic Engineering, Madison, WI, USA, 1992.
- [4] D. R. Hardoon, S. R. Szedmak, and J. R. Shawe-taylor, "Canonical correlation analysis: an overview with application to learning methods," *Neural Computation*, vol. 16, no. 12, pp. 2664-2639, 2004.
- [5] G. H. Golub and C. F. V. Loan, *Matrix Computations (Johns Hopkins Studies in Mathematical Sciences)*, Johns Hopkins University Press; 3rd edition, 1996.
- [6] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 55, pp. 2323-2326, 2000.
- [7] A. J. Ferreira and M. A. T. Figueiredo, "An unsupervised approach to feature discretization and selection," *Pattern Recognition*, vol. 45, no. 9, pp. 3048-3060, 2012.
- [8] C. Lai, M. J. T. Reinders, and L. Wessels, "Random subspace method for multivariate feature selection," *Pattern Recognition Letters*, vol. 27, no. 10, pp. 1067-1076, 2006.
- [9] A. E. Akadi, A. E. Ouardighi, and D. Aboutajdine, "A powerful feature selection approach based on Mutual Information," *International Journal of Computer Science and Network Security*, vol. 8, no. 4, pp. 116-121, 2008.
- [10] L. Yu and H. Liu, "Feature selection for high-dimensional data: a fast correlation-based filter solution," in *20th International Conference on Machine Learning*, vol. 2, pp. 856-863, 2003.
- [11] H. Uğuz, "A two-stage feature selection method for text categorization by using information gain, principal component analysis and genetic algorithm," *Knowledge-Based Systems*, vol. 24, no. 7, pp. 1024-1033, 2011.
- [12] J. Yang, Y. Liu, Z. Liu, X. Zhu, and X. Zhang, "A new feature selection algorithm based on binomial hypothesis testing for spam filtering," *Knowledge-Based Systems*, vol. 24, no. 6, pp. 904-914, 2011.
- [13] C.-M. Chen, H.-M. Lee, and C.-C. Tan, "An intelligent web-page classifier with fair feature-subset selection," *Engineering Applications of Artificial Intelligence*, vol. 19, pp. 967-978, Vancouver, BC, Canada, 2006.
- [14] H. R. Kanan and K. Faez, "An improved feature selection method based on ant colony optimization (ACO) evaluated on face recognition system," *Applied Mathematics and Computation*, vol. 205, no. 2, pp. 716-725, 2008.
- [15] Z. Yan and C. Yuan, "Ant colony optimization for feature selection in face recognition," in *Biometric Authentication*. vol. 3072, pp. 221-226, ed: Springer Berlin Heidelberg, 2004.
- [16] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, "Gene selection for cancer classification using support vector



شکل ۱۱: رتبه الگوریتم‌های انتخاب ویژگی مختلف به ازای داده‌های با سایز متوسط



شکل ۱۲: رتبه الگوریتم‌های انتخاب ویژگی مختلف به ازای داده‌های با سایز بزرگ

بر اساس این نمودار، رتبه الگوریتم پیشنهادی به ازای همه دسته‌بندها از سایر الگوریتم‌ها کمتر است. همچنین بر اساس نمودار آورده شده در شکل ۱۲ که مربوط به داده‌های حجیم است، رتبه الگوریتم انتخاب ویژگی پیشنهادی به ازای دسته‌بند KNN با اختلاف ۰/۲۹ بعد از الگوریتم MIFS-ND و به ازای دسته‌بند DT از همه الگوریتم‌های دیگر کم‌تر است.

۶- نتیجه‌گیری

در این مقاله یک روش انتخاب ویژگی از نوع فیلتر و بر اساس ترکیب نظریه بهینه پارتو و اطلاعات متقابل شرطی بنام CJMIM با دو تابع هدف جدید ارائه شد. این الگوریتم به صورت حریصانه در هر مرحله از میان ویژگی‌های کاندید ویژگی‌ای را انتخاب می‌کند که دارای حداقل افزونگی با ویژگی‌های قبلاً انتخاب شده و حداکثر ارتباط با برچسب کلاس باشد. به منظور نشان دادن کارایی الگوریتم پیشنهادی از ۱۲ مجموعه داده آموزشی معروف و پرکاربرد استفاده کردیم. دقت دسته‌بندهای KNN و DT ملاک مقایسه الگوریتم‌ها با شش الگوریتم انتخاب ویژگی معروف از جمله MIFS، MRMR، MIFS-ND، MOEA/D، ECFS و infFS قرار گرفت. بر اساس نتایج به دست آمده به طور کلی می‌توان گفت الگوریتم پیشنهادی در اکثر موارد بهتر از سایر الگوریتم‌ها عمل می‌کند.

- [34] Y. Zhang, A. Yang, C. Xiong, T. Wang, and Z. Zhang, "Feature selection using data envelopment analysis," *Knowledge-Based Systems*, vol. 64, pp. 70-80, 2014.
- [35] N. Kwak and C.-H. Choi, "Input feature selection for classification problems," *IEEE Transactions on Neural Networks*, vol. 13, no. 1, pp. 143-159, 2002.
- [36] M. T. Pablo A. Estévez and J. M. Zurada, "Normalized mutual information feature selection," presented at the *IEEE Transactions on Neural Networks*, vol. 20, no. 2, 2009.
- [37] N. Hoque, D. K. Bhattacharyya, and J. K. Kalita, "MIFS-ND: A mutual information-based feature selection method," *Expert Systems with Applications*, vol. 41, no. 14, pp. 6371-6385, 2014.
- [38] P. E. Meyer, C. Schretter, and G. Bontempi, "Information-theoretic feature selection in microarray data using variable complementarity," *IEEE Journal of Selected Topics in Signal Processing*, vol. 2, no. 2, pp. 261 - 274, 2008.
- [39] H. H. Yang and J. Moody, "Feature selection based on joint mutual information," In *Proceedings of International ICSC Symposium on Advances in Intelligent Data Analysis*, pp. 22-25, 1999.
- [40] M. Bannasar, Y. Hicks, and R. Setchi, "Feature selection using joint mutual information maximisation," *Expert Systems with Applications*, vol. 42, no. 22, pp. 8520-8532, 2015.
- [41] J. A. T. Thomas M. Cover, *Elements of Information Theory*, Wiley Series in Telecommunications and Signal Processing, Wiley-Interscience, 2006.
- [42] J. Biesiada and W. Duch, "Feature selection for high-dimensional data: a pearson redundancy based filter," *Computer Recognition Systems*, vol. 45, no. 2, pp. 242-249, 2007.
- [43] M. Haindl, P. Somol, D. Ververidis, and C. Kotropoulos, "Feature selection based on mutual correlation," *Pattern Recognition, Image Analysis and Applications*, vol. 4225, pp. 569-577, 2006.
- [44] G. Brown, A. Pocock, M.-J. Zhao, and M. Luján, "Conditional likelihood maximisation: A unifying framework for information theoretic feature selection," *The Journal of Machine Learning Research*, vol. 13, no. 1, pp. 27-66, 2012.
- [45] J. ee and D. Kim, "Fast multi-label feature selection based on information -theoretic feature ranking," *Pattern Recognition*, vol. 48, no. 9, pp. 2761-2771, 2015.
- [46] G. Roffo and S. Melzi, "Features selection via eigenvector centrality," In: *Proceedings of New Frontiers in Mining Complex Patterns (NFMCP 2016)*, 2016.
- [47] G. Roffo and S. Melzi and M. Cristani, "Infinite Feature Selection," *IEEE International Conference on Computer Vision (ICCV)*, pp. 4202-4210, Santiago, Chile, 2015.
- [48] K. Deb, S. Agrawal, A. Pratap, T. Meyarivan, "A fast elitist non-dominated sorting genetic algorithm for multi-objective optimization: Nsga-ii," *Parallel Problem Solving from Nature PPSN VI*, vol. 1917, pp. 849-858, Springer, Berlin, Heidelberg, 2000.
- [49] شیما کاشف و حسین نظام‌آبادی‌پور، «یک روش ترکیبی برای یافتن زیرمجموعه ویژگی مؤثر در داده‌های چند برچسبی»، دوره ۴۸، شماره ۳ - شماره پیاپی ۸۵، صفحه ۱۳۲۷-۱۳۳۸، ۱۳۹۷.
- [17] H. Yu, G. Gu, H. Liu, J. Shen, and J. Zhao, "A modified ant colony optimization algorithm for tumor marker gene selection," *Genomics, Proteomics & Bioinformatics*, vol. 7, no. 4, pp. 200-208, 2009.
- [18] A. Zibakhsh and M. S. Abadeh, "Gene selection for cancer tumor detection using a novel memetic algorithm with a multi-view fitness function," *Engineering Applications of Artificial Intelligence*, vol. 26, no. 4, pp. 1274-1281, 2013.
- [19] C.-L. Huang and C.-Y. Tsai, "A hybrid SOFM-SVR with a filter-based feature selection for stock market forecasting," *Expert Systems with Applications*, vol. 36, no. 2, pp. 1529-1539, 2009.
- [20] Y. Marinakis, M. Marinaki, M. Doumpos, and C. Zopounidis, "Ant colony and particle swarm optimization for financial classification problems," *Expert Systems with Applications*, vol. 36, no. 7, pp. 10604-10611, 2009.
- [21] A. Kuri-Morales and F. Rodríguez-Erazo, "A search space reduction methodology for data mining in large databases," *Engineering Applications of Artificial Intelligence*, vol. 22, no. 1, pp. 57-65, 2009.
- [22] P. M. Narendra and K. Fukunaga, "A branch and bound algorithm for feature subset selection," *IEEE Transactions on Computers* vol. 26, no. 9, pp. 917-922, 1977.
- [23] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 8, pp. 1226-1238, 2005.
- [24] R. Meiri and J. Zahavi, "Using simulated annealing to optimize the feature selection problem in marketing applications," *European Journal of Operational Research*, vol. 171, no. 3, pp. 842-858, 2006.
- [25] H. Liu and H. Motoda, *Computational Methods of Feature Selection*: Chapman & Hall/CRC, 2007.
- [26] I. A. Gheyas and L. S. Smith, "Feature subset selection in large dimensionality domains," *Pattern Recognition*, vol. 43, no. 1, pp. 5-13, 2010.
- [27] Y. Saeyns, I. Inza, and P. Larrañaga, "A review of feature selection techniques in bioinformatics," *Bioinformatics*, vol. 23, no. 19, pp. 2507-2517, 2007.
- [28] L. TH, L. HT, and s. KC, "Implementing the fisher's discriminant ratio in a k-means clustering algorithm for feature selection and data set trimming," *Journal of Chemical Information and Computer Sciences*, vol. 44, no. 1, pp. 76-87, 2004.
- [29] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, wiley interscience, 2001.
- [30] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *Journal of Machine Learning Research*, vol. 3, pp. 1157-1182, 2003.
- [31] R. Battiti, "Using mutual information for selecting features in supervised neural net learning," *IEEE Transactions on Neural Networks*, vol. 5, no. 4, pp. 537-550, 1994.
- [32] K. Kira and L. A. Rendell, "A practical approach to feature selection," *Proceedings of the ninth international workshop on Machine learning*, pp. 249-256, 1992.
- [33] J. L. F. W. C. D. Y. Qian, "A group incremental approach to feature selection applying rough set technique," *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, pp. 294-308, 2014.

- Transactions on Evolutionary Computation, vol. 11, no. 6, pp. 712 – 731, 2007 .
- [۵۳] فاطمه علیقارداشی و محمد علی زارع چاهوکی، «تأثیر ترکیب روش‌های انتخاب ویژگی فیلتر و بسته بندی در بهبود پیش بینی اشکال نرم افزار»، مجله مهندسی برق دانشگاه تبریز، دوره ۴۷، شماره ۱، صفحات ۱۸۳ تا ۱۹۵، ۱۳۹۶.
- [50] X. Zhang, X. Liu, Y. Yang, “A fast feature selection algorithm by accelerating computation of fuzzy rough set-based information entropy”, Entropy, vol. 20, issue 10, pp. 788, 2018.
- [51] F. Li, D. Miao, W. Pedrycz, “Granular multi-label feature selection based on mutual information”, Pattern Recognition, vol. 67, pp 410- 423, 2017.
- [52] Q. Zhang, H. Li, “MOEA/D: A Multiobjective Evolutionary Algorithm Based on Decomposition”, IEEE

⁶ Mutual correlation

⁷ Eigenvector Centrality

⁸ geometric series of an adjacency matrix

¹ Curse of Dimensionality

² Redundant

³ Irrelevant

⁴ Symmetrical Uncertainty

⁵ Random subspace method