

یک روش ترکیبی برای یافتن زیرمجموعه ویژگی مؤثر در داده‌های چند برچسبی

شیما کاشف^۱، دانشجوی دکتری؛ حسین نظام‌آبادی‌پور^۲، استاد

۱- دانشکده فنی مهندسی - دانشگاه شهید باهنر کرمان - کرمان - ایران - shkashef@eng.uk.ac.ir

۲- دانشکده فنی مهندسی - دانشگاه شهید باهنر کرمان - کرمان - ایران - nezam@uk.ac.ir

چکیده: داده‌های چند برچسبی به داده‌هایی گفته می‌شود که در آن بر خلاف داده‌های تک برچسبی، هر نمونه می‌تواند متعلق به چند کلاس باشد. در سال‌های اخیر، به دلیل رشد روز افزون کاربردهای این داده‌ها، طبقه‌بندی داده‌های چند برچسبی توجه بسیاری از محققان را به خود جلب کرده است. مشابه طبقه‌بندی داده‌های تک برچسبی، در داده‌های چند برچسبی نیز حذف ویژگی‌های زائد و تکراری می‌تواند تأثیر زیادی در بهبود عملکرد طبقه‌بند داشته باشد. در این مقاله، یک روش ترکیبی برای انتخاب ویژگی در داده‌های چند برچسبی ارائه شده است. روش پیشنهادی بر پایه ترکیب یک روش فیلتری و یک روش پیچشی است که در روش پیچشی از الگوریتم‌های فرا ابتکاری استفاده شده است. از آنجا که معمولاً تعداد ویژگی‌های داده‌های چند برچسبی زیاد است، استفاده مستقیم از روش‌های جستجو، برای کشف زیرمجموعه ویژگی بهینه، هزینه محاسباتی بالایی دارد و ممکن است با شکست روبه رو شود. از این رو، ابتدا با استفاده از یک روش فیلتری، ویژگی‌های نامرتب با کلاس‌ها حذف می‌شوند. سپس، از الگوریتم‌های تکاملی برای انتخاب برجسته‌ترین ویژگی‌ها استفاده می‌شود. در بخش آزمایش‌ها، تعداد قابل توجهی از الگوریتم‌های فرا ابتکاری مشهور بکار گرفته شده و جایگزین روش پیچشی در سامانه پیشنهادی شده است. نتایج به دست آمده نشان می‌دهند که روش پیشنهادی در برابر سایر روش‌های مورد مقایسه، دقت بالاتری دارند و در مواردی که دست‌یابی به دقت بالاتر، اهمیت بیشتری نسبت به زمان داشته باشد، استفاده از این روش مناسب‌تر است.

واژه‌های کلیدی: داده‌های چند برچسبی، انتخاب ویژگی، روش‌های ترکیبی، روش‌های فیلتری، روش‌های پیچشی، الگوریتم‌های تکاملی.

A Hybrid Method to Find Effective Subset of Features in Multi-label Datasets

S. Kashef¹, PhD Student; H. Nezamabadi-pour², Professor

1- Department of Electrical Engineering, Shahid Bahonar University of Kerman, Kerman, Iran, Email: shkashef@eng.uk.ac.ir

2- Department of Electrical Engineering, Shahid Bahonar University of Kerman, Kerman, Iran, Email: nezam@uk.ac.ir

Abstract: In multi-label data, each instance is associated with a set of labels, instead of one label. Due to the increasing number of modern applications associated with multi-label data, multi-label classification has gained significant attention during recent years. As in single-label data, eliminating redundant and/or irrelevant features plays an important role in improving classification performance. In this paper, a hybrid method for multi-label feature selection problem based on combining filter and wrapper methods is proposed, where meta-heuristic algorithms are employed as the wrapper method. Since, the number of features in multi-label data is usually high, solely employing search algorithms for finding the optimal feature subsets has high computational burden, and is very possible to fail. Hence, irrelevant features are first detected and removed by a filter method. Then, salient features are found among the remained features by the help of meta-heuristic algorithms. A significant number of well-known meta-heuristic algorithms are employed as the wrapper method, in the proposed system. Experiments show that the proposed method obtains better classification results, compared to other algorithms.

Keywords: Multi-label dataset, feature selection, hybrid methods, filter methods, wrapper methods, meta-heuristic algorithms.

تاریخ ارسال مقاله: ۱۳۹۵/۱۰/۱۲

تاریخ اصلاح مقاله: ۱۳۹۵/۱۲/۱۳ و ۱۳۹۶/۰۴/۰۹

تاریخ پذیرش مقاله: ۱۳۹۶/۰۵/۰۸

نام نویسنده مسئول: حسین نظام‌آبادی‌پور

نشانی نویسنده مسئول: ایران - کرمان - انتهای بلوار ۲۲ بهمن - دانشگاه شهید باهنر کرمان - دانشکده مهندسی برق.

۱- مقدمه

روش‌های فیلتر و پیچشی، دو روش مکمل هم‌اند؛ بنابراین، روش‌های ترکیبی با به کارگیری مزیت‌های هر دو روش، قادرند جواب‌های بهتری تولید کند. این ویژگی مهم، توجه بسیاری از محققان را در دهه گذشته به خود جلب کرده است [۱۰].

از بین انبوه روش‌های پیچشی، الگوریتم‌های بهینه‌سازی فرا ابتکاری نتایج بهتری داشته و بیشتر مورد توجه محققان قرار گرفتند. الگوریتم‌های جستجوی فرا ابتکاری الگوریتم‌هایی هستند که با الهام از فرایندهای فیزیکی و زیستی و طبیعی به وجود آمده‌اند و غالب آن‌ها به صورت جمعیتی عمل می‌کنند. روش‌های جستجوی ابتکاری بر خلاف روش‌های کلاسیک بر مبنای تصادف عمل کرده و جستجوی فضا را به صورت موازی انجام می‌دهند. تفاوت دیگر آن‌ها در استفاده نکردن از اطلاعات گرادیان فضا است. این نوع روش‌ها تنها از تابع برازندگی برای هدایت جستجو استفاده می‌کنند؛ اما به خاطر دارا بودن هوشمندی از نوع هوشمندی جمعی، قادر به کشف جواب می‌باشند.

در این مقاله، یک روش ترکیبی برای حل مسأله انتخاب ویژگی از داده‌های چند برجسی ارائه می‌شود. با توجه به کفایت الگوریتم‌های تکاملی در حل مسأله انتخاب ویژگی در داده‌های تک برجسی، در این مقاله قصد داریم از آنها برای ایجاد روش پیچشی استفاده کرده و به بررسی عملکرد این الگوریتم‌ها در انتخاب ویژگی در داده‌های چند برجسی بپردازیم. از این رو عملکرد شش الگوریتم شامل، ABACO [۱۰]، BACO [۱۱]، BGSa [۱۲]، CatfishPSO [۱۳]، AC0 و GA مقایسه می‌شود. شایان ذکر است که استفاده از الگوریتم‌های فرا ابتکاری برای انتخاب ویژگی داده‌های چند برجسی بسیار محدود است. یک دلیل مهم برای عدم استفاده از این الگوریتم‌ها در انتخاب ویژگی داده‌های چند برجسی، تعداد زیاد ویژگی‌های داده‌های چند برجسی در مقابل داده‌های تک برجسی می‌باشد. به همین دلیل، استفاده از این الگوریتم‌ها، به تنهایی، ممکن است منجر به عدم موفقیت در یافتن زیرمجموعه ویژگی‌های بهینه شود. از این رو، در روش پیشنهادی ابتدا با استفاده از یک روش فیلتری، تعداد قابل توجهی از ویژگی‌های نامرتب و افزونه حذف شده، سپس از این الگوریتم‌ها برای یافتن زیرمجموعه ویژگی بهینه استفاده می‌شود.

ادامه این مقاله این گونه سازماندهی شده است که در بخش دوم مفاهیم پایه مربوط به داده‌های چند برجسی تشریح می‌شود. در بخش سوم، کارهای انجام شده در زمینه انتخاب ویژگی داده‌های چند برجسی مرور خواهد شد. در بخش چهارم، روش پیشنهادی ترکیبی برای انتخاب ویژگی در داده‌های چند برجسی بیان می‌شود. مقایسه کارایی روش‌های پیشنهادی روی چند مجموعه داده استاندارد در بخش پنجم آورده شده است. در نهایت، در بخش ششم، مقاله جمع بندی می‌شود.

یادگیری با نظارت سنتی یکی از بحث‌های پرطرفدار در حوزه یادگیری ماشینی است که در آن هر نمونه با تعدادی ویژگی و یک برجسب y از مجموعه برجسب‌های L مشخص می‌شود. با این حال در مسائل دنیای واقعی، هر نمونه ممکن است به یک مجموعه برجسب $V_i \subset L$ تعلق داشته باشد. به طور مثال، یک تصویر ممکن است به طور همزمان با دو برجسب "ساحل" و "غروب خورشید"، حاشیه نویسی شود. یعنی بر خلاف یادگیری با نظارت سنتی، در یادگیری چند برجسی^۱، علیرغم اینکه هر شیء با یک نمونه نشان داده می‌شود، دارای مجموعه‌ای از برجسب‌ها به جای یک برجسب است. در سال‌های اخیر، تعداد رو به افزایشی از کاربردهای مدرن، نظیر طبقه‌بندی متن [۱]، احساسات برخواسته از موسیقی [۲]، برجسب زنی معنایی تصاویر^۲ [۳] و ویدئوها [۴]، طبقه بندی تابع پروتئین [۵] و ژن‌ها [۶]، که دارای داده‌های چند برجسی هستند، پدید آمده است.

به طور کلی، داده‌های چند برجسی دارای تعداد زیادی ویژگی هستند که نمونه‌ها را توصیف می‌کنند. ویژگی‌های نامرتب، تکراری و نویزی تأثیر بسیار بدی روی کارایی الگوریتم‌های یادگیری دارند [۷]. به علاوه، یافتن مقادیر بعضی از ویژگی‌ها به قدری هزینه بر است که تحمل مقداری افزایش در خطای طبقه‌بندی در مقابل حذف آن ویژگی ترجیح داده می‌شود. اگرچه معمولاً دقت طبقه‌بند بعد از انتخاب ویژگی بهبود می‌یابد.

تا کنون برای انتخاب ویژگی روش‌های متنوعی استفاده شده است که به طور کلی می‌توان آن‌ها را به سه دسته اصلی تقسیم کرد: روش‌های فیلتری^۳، روش‌های پیچشی^۴ و روش‌های ترکیبی^۵. روش‌های فیلتری، مستقل از الگوریتم‌های یادگیری عمل کرده و با استفاده از معیارهایی خاص، برای هر ویژگی یک اعتبار محاسبه می‌کنند [۸]. سپس، ویژگی‌ها بر اساس رتبه‌هایشان مرتب شده و ویژگی‌هایی که رتبه آن‌ها کمتر از یک مقدار آستانه از پیش تعیین شده است، حذف می‌گردند. این روش‌ها از سرعت بالایی برخوردارند و برای داده‌های با ابعاد زیاد مناسب‌اند؛ اما دقت آن‌ها در مقایسه با روش‌های پیچشی پایین‌تر است. تاکنون روش‌های فیلتر زیادی نظیر F-score، اطلاعات متقابل^۶، بهره اطلاعاتی^۷، همبستگی^۸ و غیره در زمینه انتخاب ویژگی مطرح شده است [۹]. در روش پیچشی، بازخوردی از الگوریتم یادگیری اعمال شده برای انتخاب ویژگی استفاده می‌شود. به عبارت دیگر، معیار انتخاب ویژگی، کارایی الگوریتم طبقه‌بندی است. بنابراین، این روش از دقت بالایی برخوردار است. اما از آنجا که برای بررسی میزان شایستگی هر زیرمجموعه، باید الگوریتم طبقه‌بند اجرا شود، سرعت کار پایین و پیچیدگی محاسباتی بالا می‌رود و این امر استفاده از این روش را در داده‌های با ابعاد زیاد محدود می‌سازد.

۲- مفاهیم پایه

برای طبقه‌بندی یک نمونه دیده نشده، BR با جستجوی برچسب‌های مثبت روی هر یک از طبقه‌بندهای دودویی، این نمونه را طبقه‌بندی می‌کند و در آخر، ترکیب نتایج هر یک از این طبقه‌بندها را به عنوان برچسب‌های این نمونه تحویل می‌دهد. مزیت طبقه‌بند BR ، سادگی آن است؛ اما عیب بزرگی که دارد این است که این طبقه‌بند قادر به کشف روابط بین برچسب‌ها نیست [۱۹].

۲-۲- Label Powerset

این روش، داده چند برچسبی را به یک مجموعه داده تک برچسبی چند کلاسه تبدیل می‌کند. برای این کار، LP هر ترکیب یکتا از برچسب‌ها در مجموعه داده چند برچسبی را به عنوان یک کلاس برای مجموعه داده چند کلاسه متناظر در نظر می‌گیرد. به عبارت دیگر، هر نمونه $(x_i, Y_i), i = 1 \dots N$ ، به یک نمونه (x_i, l_i) تبدیل می‌شود که l_i برچسب ساخته شده‌ای است که بیانگر یک زیرمجموعه برچسب مجزا و یکتا است. بنابراین، LP بر خلاف BR ، وابستگی بین برچسب‌ها را در نظر می‌گیرد. با این وجود، از آنجا که تعداد کلاس‌های مجموعه داده چند کلاسه ساخته شده، متناسب با تعداد زیرمجموعه برچسب‌های یکتا در D می‌باشد، مشکل اصلی این روش این است که بعضی از کلاس‌ها در مجموعه داده چند کلاسه ساخته شده ممکن است متعلق به تعداد بسیار کمی از نمونه‌ها باشند. در نتیجه، یک مجموعه داده نامتوازن^{۱۲} ساخته شده است [۱۴].

۳- مروری بر کارهای انجام شده

همانند داده‌های تک برچسبی، انتخاب ویژگی در داده‌های چند برچسبی هم به سه دسته اصلی با همان تعاریف تقسیم می‌شود. در اکثر روش‌های انتخاب ویژگی چند برچسبی، ابتدا با استفاده از یک روش تغییر شکل مسئله نظیر BR یا LP ، داده چند برچسبی را به داده تک برچسبی تغییر می‌دهند. سپس، هر روش انتخاب ویژگی تک برچسبی برای انتخاب ویژگی‌های برجسته قابل استفاده است. مرجع [۱۴] بعد از اعمال دو روش BR و LP برای دستیابی به داده تک برچسبی، از دو روش فیلتری بهره اطلاعات و ReliefF برای انتخاب ویژگی استفاده کرده است. در مرجع [۲۰] نویسندگان از یک روش تغییر شکل مسئله هرس شده^{۱۳} (ppt) که قبلاً در مرجع [۲۱] معرفی شده است، برای تغییر مجموعه داده چند برچسبی به مجموعه داده تک برچسبی استفاده کرده‌اند. سپس، یک روش انتخاب ویژگی حریص^{۱۴} مبتنی بر اطلاعات متقابل چند بعدی^{۱۵} استفاده شده است. روش مشابهی در مرجع [۲۲] پیشنهاد شده است که با استفاده از ppt داده چند برچسبی را به داده تک برچسبی تبدیل می‌کند و سپس از الگوریتم ReliefF برای وزندهی به ویژگی‌ها استفاده می‌کند. از سوی دیگر، برخی الگوریتم‌های انتخاب ویژگی چند برچسبی به طور مستقیم با مجموعه داده چند برچسبی کار می‌کنند. به طور مثال، در مرجع [۲۳] یک روش انتخاب ویژگی چند برچسبی مبتنی بر اطلاعات متقابل چند بعدی پیشنهاد شده است که زیرمجموعه ویژگی‌های برجسته را با بیشینه کردن اطلاعات متقابل چند

فرض کنید $\mathcal{X} = \mathbb{R}^M$ (or \mathbb{Z}^M) نشان دهنده فضای نمونه M بعدی و $L = \{y_1, y_2, \dots, y_q\}$ نشان دهنده فضای برچسب‌ها با q برچسب ممکن باشد. در یادگیری چند برچسبی، هدف، یادگیری یک تابع $h: \mathcal{X} \rightarrow 2^q$ از مجموعه داده آموزش $D = \{(x_i, Y_i), i = 1 \dots N\}$ با N نمونه است، که هر نمونه دارای یک بردار ویژگی $x_i = (x_{i1}, x_{i2}, \dots, x_{iM})$ است که با M ویژگی توصیف می‌شود. شکل ۱ این نمایش را نشان می‌دهد. برای هر نمونه دیده نشده $x \in \mathcal{X}$ ، طبقه‌بند چند برچسبی $h(\cdot)$ از مجموعه برچسب‌های مناسب برای x را پیش بینی می‌کند $L \rightarrow h(x)$.

روش‌های یادگیری چند برچسبی به دو دسته کلی تقسیم می‌شوند: تغییر شکل مسئله^{۱۰} و انطباق الگوریتم^{۱۱} [۱۴]. در روش اول، مسئله یادگیری چند برچسبی به یک یا چند مسئله تک برچسبی تغییر می‌کند و سپس هر الگوریتم یادگیری تک برچسبی می‌تواند مورد استفاده قرار گیرد. دو روش Binary Relevance (BR) و Label Powerset (LP) از این دسته‌اند که در ادامه توضیح داده می‌شوند. در روش دوم، الگوریتم‌های یادگیری به گونه‌ای طراحی می‌شوند که بتوانند به طور مستقیم با داده‌های چند برچسبی کار کنند. الگوریتم‌هایی نظیر بیز ساده چند برچسبی^{۱۱} [۱۵]، الگوریتم یادگیری تنبل MLKNN [۱۶] و روش درخت تصمیم چند برچسبی [۱۷] از این دسته‌اند.

X	Y			
	y_1	y_2	\dots	y_q
x_{11} x_{12} \dots x_{1M}	0	1	1	0
x_{21} x_{22} \dots x_{2M}	1	1	0	1
\vdots \vdots \ddots \vdots	\vdots	\vdots	\vdots	\vdots
x_{N1} x_{N2} \dots x_{NM}	1	0	1	0

شکل ۱: داده چند برچسبی

۲-۱- Binary Relevance

روش BR یک راهبرد تغییر شکل مسئله است که در آن، مسئله طبقه‌بندی چند برچسبی به $(|L|)$ مسئله مجزای طبقه‌بندی تک برچسبی تبدیل می‌شود. به عبارت دیگر، مجموعه داده اصلی، به q مجموعه داده برچسب هستند. درایه‌های هر بردار برچسب، با توجه به تعلق یا عدم تعلق هر نمونه به آن برچسب به صورت y_i و $-y_i$ می‌باشد که به ترتیب نمونه مثبت و منفی برای آن برچسب خوانده می‌شوند [۱۸]. به طور مثال، در یک مجموعه داده ۴ برچسبی، اگر یک نمونه دارای دو برچسب $\{y_2, y_4\}$ باشد، درایه متناظر آن در مجموعه داده‌های دوم و چهارم برابر با y_2 و y_4 و در مجموعه داده‌های اول و سوم برابر با $-y_1$ و $-y_3$ است.

برچسبی به داده‌های تک برچسبی، نوبت به مرحله انتخاب ویژگی می‌رسد که خود، شامل دو مرحله است. ابتدا، انتخاب ویژگی با استفاده از یک روش فیلتری روی داده‌های تک برچسبی اجرا می‌شود. در این مرحله تعدادی از ویژگی‌های نامرتب با کلاس‌ها حذف می‌شوند. سپس، ویژگی‌های انتخاب شده از این مرحله روی داده چند برچسبی اصلی اعمال شده، که منجر به ایجاد یک داده چند برچسبی کاهش بعد یافته می‌شود. اکنون، با استفاده از الگوریتم‌های فرا ابتکاری به عنوان روش پیچشی و یک طبقه‌بند چند برچسبی، بهترین زیرمجموعه ویژگی از میان این ویژگی‌ها انتخاب می‌شود. شرح دقیق روش پیشنهادی در ادامه آورده شده است.

ابتدا با استفاده از روش BR، داده چند برچسبی به q داده تک برچسبی تبدیل می‌شود. سپس، نوبت به بلوک انتخاب ویژگی می‌رسد. با توجه به زیاد بودن تعداد ویژگی‌های داده‌های چند برچسبی، ابتدا از یک روش فیلتری برای حذف ویژگی‌های نامرتب با کلاس‌ها، استفاده می‌شود. برای این کار، روش فیلتری مورد نظر روی هر یک از q مجموعه داده اعمال شده که منجر به انتخاب یک سری از ویژگی‌ها، در هر یک از این مجموعه‌ها می‌شود. این ساختار در شکل ۳ آورده شده است. در حقیقت، شکل ۳ بلوک انتخاب ویژگی با استفاده از روش فیلتری به صورت تک برچسبی را نشان می‌دهد که سومین بلوک در شکل ۲ است. همانطور که می‌دانیم، ویژگی‌های مؤثر، ویژگی‌هایی هستند که دارای دو خصلت مهم باشند. اولاً بتوانند به خوبی کلاس‌ها را از یکدیگر تمیز دهند؛ به عبارتی با کلاس‌ها مرتبط باشند. ثانیاً، اینکه این ویژگی‌ها تکراری نباشند؛ زیرا هدف، رسیدن به دقت طبقه‌بندی قابل قبول با داشتن کمترین تعداد ویژگی است. در داده‌های چند برچسبی، تشخیص ویژگی‌های تکراری کار آسانی نیست. به طور مثال، ممکن است دو ویژگی بر اساس معیارهای متداول در داده‌های تک برچسبی، تکراری محسوب شوند. اما با بررسی دقیق‌تر مشاهده می‌شود که هر یک از دو ویژگی با برچسب‌های متفاوتی مرتبط هستند. به عبارت دیگر، برچسب‌هایی را که یک ویژگی پوشش می‌دهد و می‌تواند به خوبی تعلق یا عدم تعلق یک نمونه به آن برچسب را تشخیص دهد، با برچسب‌های تحت پوشش ویژگی دوم متفاوت است. در نتیجه نمی‌توان به سادگی ویژگی‌ها را تکراری خواند و آن‌ها را حذف کرد [۳۳]. تدبیری که در این مقاله اندیشیده شده است، به این صورت است که ابتدا ویژگی‌های نامرتب با کلاس‌ها حذف شوند.

بعدی بین ویژگی‌ها و برچسب‌ها پیدا می‌کند. همچنین، روش‌های انتخاب ویژگی چند برچسبی متفاوتی که مبتنی بر روش فیلتری تک برچسبی Relief-F می‌باشند، در مراجع [۲۲، ۲۴، ۲۵] معرفی شده‌اند. مرجع [۲۶] روش انتخاب ویژگی چند برچسبی مبتنی بر الگوریتم ممتیک را پیشنهاد کرده که از فرآیند ممتیک برای پیدا کردن ویژگی‌های برجسته از میان ویژگی‌هایی که با جستجوی وراثتی انتخاب شده بودند، استفاده می‌کند. در مرجع [۲۷] روش انتخاب ویژگی چند برچسبی افزایشی^{۱۶} مبتنی بر وابستگی بیشینه و تکراری بودن کمینه^{۱۷} پیشنهاد شده است که از روش معروف فیلتری mRMR [۲۸] برای داده‌های تک برچسبی، الهام گرفته است.

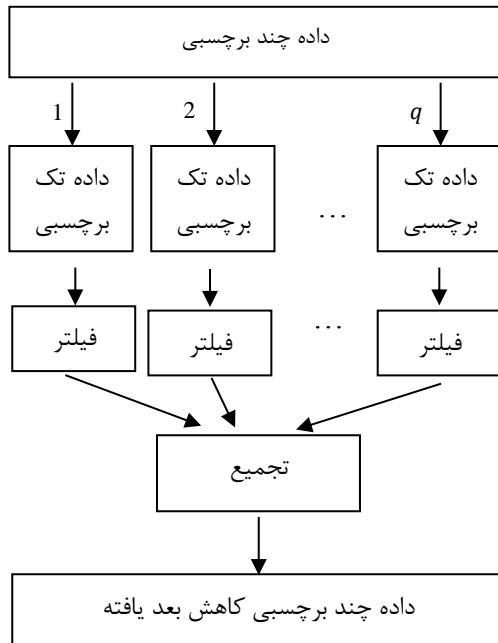
مرجع [۲۹] روشی بر مبنای بهینه‌سازی چند هدفه برای انتخاب ویژگی داده‌های چند برچسبی ارائه داده است. نویسندگان، با استفاده از الگوریتم چند هدفه NSGA-II تلاش کردند زیرمجموعه ویژگی را بیابند که معیار میانگین صحت را بیشینه و معیار افت همینگ را کمینه کند. مرجع [۳۰] روشی به نام LCFS پیشنهاد داده است که با استفاده از روابط بین برچسب‌های اصلی، برچسب‌های جدید می‌سازد. سپس، برچسب‌های ساخته شده به مجموعه داده اصلی اضافه می‌شوند و با استفاده از روش IG-BR انتخاب ویژگی از بین این مجموعه داده بسط داده شده انجام می‌گیرد. در مرجع [۳۱] ابتدا به این مطلب اشاره شده است که در یک مجموعه با n ویژگی، تعداد کل زیرمجموعه ویژگی‌های موجود برابر با 2^n حالت است و احتمال به دام افتادن روش‌های جستجوی حریصانه در بهینه‌های محلی وجود دارد. برای حل این مشکل، در این مقاله تابع امتیازی بر مبنای اطلاعات متقابل، پیشنهاد و یک روش بهینه‌سازی عددی برای جلوگیری از به دام افتادن در بهینه‌های محلی ارائه شده است. در مرجع [۳۲] روشی برای انتخاب دقیق‌تر ویژگی‌های مرتبط پیشنهاد شده است. نویسندگان با تبیین ضعف روش‌های پیشین در آشکارسازی ویژگی‌های مرتبط در مجموعه داده‌های با تعداد زیاد برچسب، روشی را برای اصلاح یکی از روش‌های پیشین در تعیین ویژگی‌های مرتبط پیشنهاد دادند.

۴- روش پیشنهادی

همانطور که قبلاً گفته شد، در این مقاله قصد داریم روشی برای اعمال انواع الگوریتم‌های فرا ابتکاری در انتخاب ویژگی داده‌های چند برچسبی ارائه کنیم، که با وجود تعداد زیاد ویژگی در این داده‌ها، قادر باشند در یک زمان قابل قبول، زیرمجموعه ویژگی بهینه را بیابند.

شکل ۲ روند کلی روش پیشنهادی در انتخاب ویژگی داده‌های چند برچسبی را نشان می‌دهد. طبق این شکل، بعد از تبدیل داده چند

را می‌طلبید، بر عهده الگوریتم‌های فرا ابتکاری گذاشته می‌شود. بنابراین، پس از فیلتر کردن ویژگی‌های نامرتبط، با استفاده از الگوریتم‌های فرا ابتکاری، بهترین ویژگی‌ها از بین ویژگی‌های باقی مانده انتخاب می‌شوند.

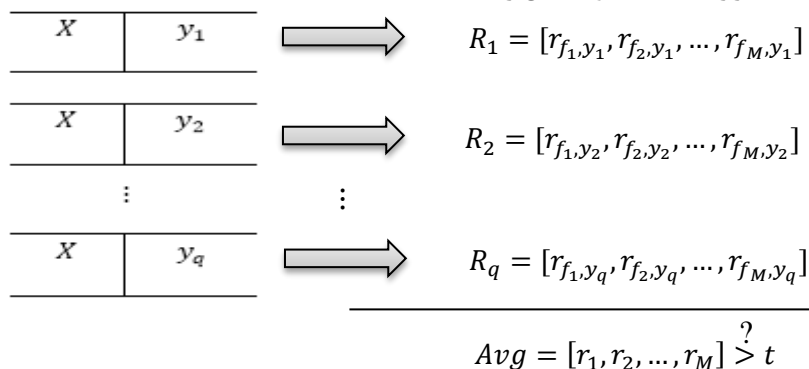


شکل ۳: ساختار روش پیشنهادی برای بلوک انتخاب ویژگی فیلتری

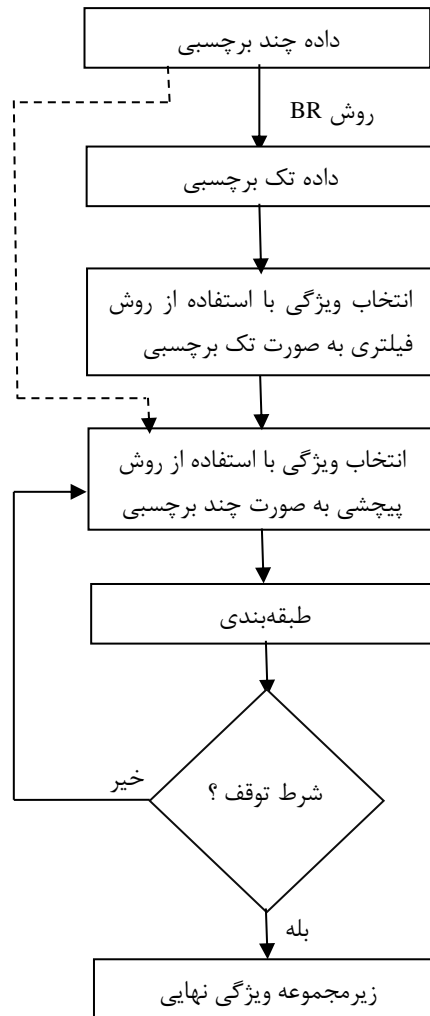
برای اعمال روش فیلتری و یافتن ویژگی‌های مرتبط با کلاس‌ها از ضریب همبستگی پیرسون^{۱۸} استفاده می‌شود. در داده‌های تک برجسی، ضریب همبستگی r_{FL} بین ویژگی F با مقادیر f و برجسب L با مقادیر l عبارتست از [۳۵]:

$$r_{FL} = \frac{\sum_i (f_i - \bar{f}_i)(l_i - \bar{l}_i)}{\sqrt{\sum_i (f_i - \bar{f}_i)^2} \sqrt{\sum_j (l_j - \bar{l}_j)^2}} \quad (1)$$

در اینجا \bar{f}_i و \bar{l}_i به ترتیب میانگین مقادیر میانگین مقادیر f_i و l_i روی N نمونه است. اگر ویژگی F با برجسب L کاملاً همبسته باشند، یعنی



شکل ۴: فرآیند انتخاب ویژگی‌های مرتبط در روش فیلتری



شکل ۴: روند کلی روش پیشنهادی در انتخاب ویژگی داده‌های چند برجسی

توجه به این نکته ضروری است که در این مرحله، از روش فیلتری استفاده می‌شود که زیاد قدرتمند نباشد و منجر به حذف تعداد زیادی از ویژگی‌ها نشود. به طور مثال، روش فیلتری مبتنی بر همبستگی سریع که در مرجع [۳۴] معرفی شده است، روشی قدرتمند در انتخاب ویژگی داده‌های تک برجسی است که تعداد ویژگی‌ها را به شدت کم می‌کند. هدف از این کار، حذف ویژگی‌های نامرتبط با برجسب‌ها است که مطمئناً زائد هستند. کشف و حذف ویژگی‌های تکراری که جستجوی دقیق‌تری

این معیار در رابطه ۲ آورده شده است. همچنین، از طبقه‌بند ML-kNN [۱۶] که نمونه چند برچسبی شده طبقه‌بند معروف kNN است، برای طبقه‌بندی استفاده می‌شود.

$$F_{measure} = \frac{2 \times Recall \times Precision}{Recall + Precision} \quad (2)$$

۵- آزمایش‌ها و نتایج

در این بخش، نتایج روش پیشنهادی در استفاده از الگوریتم‌های تکاملی برای انتخاب ویژگی داده‌های چند برچسبی روی ۵ مجموعه داده با اندازه‌ها و کاربردهای متفاوت بررسی می‌شوند. همچنین، برای مشاهده قدرت این الگوریتم‌ها در انتخاب ویژگی، نتایج چند روش معروف در انتخاب ویژگی در داده چند برچسبی نیز گزارش شده است.

۵-۱- پایگاه داده

برای بررسی نتایج، ۵ مجموعه داده از کاربردهای مختلف، از پایگاه داده مولان^{۲۱} تهیه شده است. جدول ۱ مشخصه‌های این مجموعه داده‌ها، شامل اسم، دامنه، تعداد نمونه‌ها، تعداد ویژگی‌ها، تعداد برچسب‌ها $|L| = q$ ، نوع ویژگی‌ها، کاردینالیته برچسب‌ها^{۲۲}، که میانگین تعداد برچسب‌های هر نمونه تعریف می‌شود و چگالی برچسب‌ها^{۲۳} که نرمالیزه شده کاردینالیته با $|L|$ است و به ترتیب با روابط ۳ و ۴ تعریف می‌شوند.

$$LC(D) = \frac{1}{|D|} \sum_{i=1}^{|D|} |Y_i| \quad (3)$$

$$LD(D) = \frac{1}{|D|} \sum_{i=1}^{|D|} \frac{|Y_i|}{|L|} \quad (4)$$

جدول ۱: مشخصه مجموعه ویژگی‌های مختلف استفاده شده

دامنه	چگالی برچسب‌ها	کاردینالیته برچسب‌ها	نوع ویژگی‌ها	تعداد برچسب‌ها	تعداد ویژگی‌ها	تعداد نمونه‌ها	مجموعه داده
موسیقی	۰/۱۱۵	۲۶/۰۴	عددی	۱۷۴	۶۸	۵۰۲	cal500
موسیقی	۰/۳۱۱	۱/۸۶۹	عددی	۶	۷۲	۵۹۳	emotions
متن	۰/۰۲۸	۱/۲۴۵	اسمی	۴۵	۱۴۴۹	۹۷۸	medical
متن	۰/۰۰۶	۳/۳۸	اسمی	۵۳	۱۰۰۱	۱۷۰۲	enron
زیست‌شناسی	۰/۳۰۳	۴/۲۳۷	عددی	۱۴	۱۰۳	۲۴۱۷	yeast

است. فرض کنید $T = \{(t_i, Y_i), i = 1, \dots, p\}$ یک مجموعه آزمون باشد که $Y_i \subseteq L$ زیرمجموعه صحیح و $Z_i \subseteq L$ زیرمجموعه پیش‌بینی شده متناظر با نمونه t_i باشد. همچنین فرض کنید $f(x, y)$ نشان دهنده امتیازی است که به برچسب y برای نمونه x داده می‌شود. این معیارها در ادامه معرفی شده‌اند [۱۹].

۵-۲- معیارهای ارزیابی

برای ارزیابی عملکرد روش‌های پیشنهادی، از معیارهای مربوط به داده‌های چند برچسبی شامل دقت^{۲۴}، افت همینگ^{۲۵}، یک-خطا^{۲۶}، پوشش^{۲۷}، افت رتبه^{۲۸}، میانگین صحت^{۲۹} و میانگین کاهش ویژگی‌ها^{۳۰} استفاده شده

• دقت

این معیار که نسبت برچسب‌های پیش بینی شده صحیح را به تعداد کل برچسب‌های پیش بینی شده و صحیح محاسبه می‌کند، به صورت زیر تعریف می‌شود:

$$accuracy(h, T) = \frac{1}{|T|} \sum_{i=1}^{|T|} \frac{|Y_i \cap Z_i|}{|Y_i \cup Z_i|} \quad (5)$$

• افت همینگ

افت همینگ درصد برچسب‌هایی که به درستی طبقه‌بندی نشده‌اند را محاسبه می‌کند؛ یعنی نمونه‌ای که به برچسب نادرست نسبت داده شده باشد یا برچسب صحیحی که به یک نمونه متعلق باشد اما پیش بینی نشده باشد.

$$Hamming Loss(h, T) = \frac{1}{|T|} \sum_{i=1}^{|T|} \frac{|Y_i \Delta Z_i|}{|L|} \quad (6)$$

Δ تفاوت متقارن^{۲۱} بین دو مجموعه است.

• یک-خطا

این معیار تعداد دفعاتی را می‌شمارد که برچسب با بالاترین درجه تعلق به یک نمونه، متعلق به آن نمونه نباشد.

$$one - error(f) = \frac{1}{|T|} \sum_{i=1}^{|T|} [|\arg \max_{y \in Y} f(x_i, y)| \notin Y_i] \quad (7)$$

• پوشش

معیار پوشش میانگین تعداد گام‌هایی را می‌شمارد که باید از لیست برچسب‌های رتبه بندی شده پایین رفت تا تمام برچسب‌های مربوط به هر نمونه پوشش داده شود.

$$coverage(f) = \frac{1}{|T|} \sum_{i=1}^{|T|} \max_{y \in Y} rank_f(x_i, y) - 1 \quad (8)$$

$rank_f(t_i, y)$ نشان دهنده رتبه y در Y است که بر اساس ترتیب نزولی به وسیله f تعیین شده است.

• افت رتبه

افت رتبه میانگین تعداد دفعاتی را می‌شمارد که یک برچسب مرتبط رتبه‌ای بالاتر از یک برچسب نامرتب گرفته باشد؛ یعنی ترتیب برچسب‌ها معکوس شده باشد.

$$rloss(f) = \frac{1}{|T|} \sum_{i=1}^{|T|} \frac{1}{|Y_i \cap \bar{Y}_i|} |\{(y', y'') | f(x_i, y') \leq f(x_i, y''), (y', y'') \in Y_i \times \bar{Y}_i\}| \quad (9)$$

• میانگین صحت

این معیار میانگین درصد برچسب‌های مرتبط را که بالای یک برچسب مرتبط خاص $Y_i \in Y$ رتبه‌بندی شده‌اند را محاسبه می‌کند.

$$avgprec(f) = \frac{1}{|T|} \sum_{i=1}^{|T|} \frac{1}{|Y_i|} \sum_{y \in Y_i} \frac{|\{y' | rank_f(x_i, y') \leq rank_f(x_i, y), y' \in Y_i\}|}{rank_f(x_i, y)} \quad (10)$$

• نرخ کاهش ویژگی‌ها^{۲۲}

برای بررسی تعداد ویژگی‌ها بعد از انتخاب ویژگی، نسبت به تعداد ویژگی‌های اصلی، پارامتری به نام F_r تعریف می‌شود که با رابطه زیر به دست می‌آید [۹]:

$$F_r = \frac{M - q}{M}$$

که M تعداد ویژگی‌های اصلی و q تعداد ویژگی‌ها بعد از انتخاب ویژگی است. طبق این فرمول هرچه مقدار F_r به ۱ نزدیک‌تر باشد، کاهش تعداد ویژگی‌ها بیشتر بوده و مطلوب‌تر است.

۵-۳- الگوریتم‌های مقایسه و تنظیم پارامترها

همانطور که قبلاً گفته شد، الگوریتم‌های فرا ابتکاری که برای حل مسأله انتخاب ویژگی در داده‌های چند برچسبی مورد آزمایش قرار گرفتند عبارتند از: الگوریتم مورچگان دودویی پیشرفته ABACO [۱۰]، الگوریتم گسسته مورچگان ACO، الگوریتم مورچگان باینری BACO [۱۱]، الگوریتم وراثتی GA الگوریتم گرانشی دودویی BGSA [۱۲] و الگوریتم جمعیت ذرات باینری با خاصیت گربه ماهی CatfishBPSO [۱۳].

در تمام الگوریتم‌های مورد مقایسه، اندازه جمعیت و تعداد تکرارها، به ترتیب برابر ۵۰ و ۲۰ در نظر گرفته شده‌اند. در پیاده سازی اقسام مختلف الگوریتم مورچه، $\tau_0 = 0.1$ (رد پای اولیه)، $\tau_{min} = 0.1$ (مینیمم رد پای)، $\tau_{max} = 6$ (ماکزیمم رد پای)، هزینه سفر (Fk) برابر عکس نرخ بازشناسی درست و پارامترهای کنترلی α و β به ترتیب برابر ۱ و ۰ در نظر گرفته شده‌اند. مقدار ρ و Q به ترتیب ۰.۵ و ۰.۳ در نظر گرفته شدند. در الگوریتم ژنتیک، همبری ۳۳ از نوع تک نقطه‌ای با احتمال ۰.۹ و جهش ۳۴ از نوع دودویی با احتمال ۰.۱ لحاظ شده‌اند. در الگوریتم جمعیت ذرات، وزن اینرسی $w=1$ و ضرایب $c_1 = c_2 = 2$ در نظر گرفته شده‌اند.

در همه الگوریتم‌های مورد مقایسه، تعداد ویژگی‌ها در حین کار الگوریتم‌ها تعیین می‌شود. تنها در الگوریتم مورچگان گسسته، ACO، این تعداد باید توسط کاربر مشخص شود. بنابراین برای ایجاد شرایط یکسان در مقایسه کارکرد الگوریتم‌ها، ابتدا الگوریتم ABACO برای ۲۰ مرتبه به صورت مستقل اجرا شد و سپس میانگین تعداد ویژگی‌هایی را که این الگوریتم برای هر مجموعه داده انتخاب کرده بود، به عنوان ورودی به ACO اعمال کردیم.

برای مقایسه عملکرد الگوریتم‌های تکاملی در یافتن زیرمجموعه ویژگی بهینه، نسبت به سایر روش‌های پیشنهادی برای حل مسأله انتخاب ویژگی در داده‌های چند برچسبی، نتایج چهار الگوریتم شامل روش‌های LP-RF، LP-IG، BR-RF و BR-IG [۱۴] نیز آورده شده است. در این الگوریتم‌ها، ابتدا از دو روش LP و BR برای تغییر داده‌های چند برچسبی به تک برچسبی استفاده می‌شود. سپس، از دو روش فیلتری بهره اطلاعات (IG) و ReliefF (RF) برای زنده‌ی ویژگی‌ها

در برخی معیارها، از جمله دقت، میزان بهبود نتایج به دست آمده به وسیله الگوریتم‌های تکاملی نسبت به چهار روش BR-IG، LP-IG، LP-RF و RF محسوس‌تر از سایر معیارها است. به طور مثال، در مجموعه داده emotions میزان بهبود معیار دقت در روش‌های هایبرید (الگوریتم‌های تکاملی) نسبت به روش‌های فیلتری در حدود ۱۵٪ است؛ در حالیکه میزان بهبود نتایج برای همین مجموعه داده در معیار افت همینگ در حدود ۱٪ است.

در جدول ۹، میانگین زمان اجرای روش‌های مورد مقایسه در ۲۰ اجرای مجزا آورده شده است. همانطور که انتظار می‌رود، روش‌های فیلتری LP-RF، LP-IG، BR-RF و BR-IG زمان پردازش کمتری نسبت به الگوریتم‌های تکاملی دارند که در دسته روش‌های هایبرید قرار می‌گیرند. زیرا در هر تکرار، از سامانه یادگیری برای ارزیابی کیفیت زیرمجموعه ویژگی انتخاب شده سؤال می‌شود. اما آنجا که فرآیند انتخاب ویژگی به صورت برون خط^{۲۶} انجام می‌شود، این مقدار تفاوت زمانی در مقابل انتخاب ویژگی‌های موثرتر و افزایش کارایی طبقه‌بند قابل تحمل است و ترجیح داده می‌شود.

۶- جمع بندی و پیشنهادها

در داده‌های چند برجسی بر خلاف داده‌های تک برجسی که هر نمونه دارای یک برجسب است، هر نمونه می‌تواند به طور همزمان متعلق به چندین برجسب باشد. با توجه به ظهور کاربردهای مدرن برای داده‌های چند برجسی، انتخاب ویژگی برای افزایش دقت طبقه‌بندی و کاهش پیچیدگی، امری ضروری است. در این مقاله، از شش الگوریتم تکاملی پر کاربرد، برای حل مسأله انتخاب ویژگی در داده‌های چند برجسی استفاده شده است. از آنجا که معمولاً تعداد ویژگی‌های داده‌های چند برجسی زیاد است، استفاده از الگوریتم‌های تکاملی برای یافتن زیرمجموعه ویژگی بهینه، پر هزینه است و عملی نمی‌باشد. برای حل این مشکل، ابتدا با استفاده از یک روش فیلتری، ویژگی‌های نامرتب با کلاس‌ها حذف شدند. سپس، با استفاده از الگوریتم‌های تکاملی زیرمجموعه ویژگی بهینه از میان ویژگی‌های باقی مانده انتخاب شدند. برای ارزیابی میزان پیشرفت نتایج، چهار روش جدید در انتخاب ویژگی داده‌های چند برجسی نیز با شرایط یکسان، روی مجموعه داده‌های مشابه اعمال شده و نتایج آن‌ها نیز گزارش شدند. با توجه به نتایج به دست آمده، میزان بهبود در تمامی معیارهای ارزیابی روی اکثر مجموعه داده‌ها، قابل توجه است.

در این مقاله، از معیار F به عنوان تابع شایستگی برای الگوریتم‌های تکاملی استفاده شده است. برای بهبود عملکرد الگوریتم در سایر معیارها، می‌توان تابع شایستگی را بر اساس معیارهای دیگر تعریف کرد. همچنین، می‌توان از روش فیلتری قوی‌تری برای حذف ویژگی‌های زائد در مرحله اول استفاده کرد تا تعداد ویژگی‌های بیشتری فیلتر شوند و کار الگوریتم‌های تکاملی برای یافتن بهترین ویژگی‌ها از میان ویژگی‌های باقی مانده سبک‌تر شود.

استفاده می‌شود. در نهایت ویژگی‌هایی که وزن بالاتر از یک مقدار آستانه دارند، به عنوان ویژگی‌های نهایی انتخاب می‌شوند. این حد آستانه طبق مرجع [۱۴] ۰/۰۱ در نظر گرفته شده است. این الگوریتم‌ها نیز ۲۰ مرتبه به طور مستقل اجرا شده و میانگین نتایج استخراج شدند.

۵-۴- نتایج

برای طبقه‌بندی از طبقه‌بند k همسایه نزدیک‌تر برای داده‌های چند برجسی (ML-kNN) استفاده شده است (k=۱۰). این طبقه‌بند، مدل بسط داده شده طبقه‌بند معروف kNN برای داده‌های تک برجسی است. به طور دقیق‌تر، برای یک طبقه‌بندی یک داده دیده نشده، ابتدا k همسایه نزدیک‌تر در مجموعه آموزش، برای این نمونه پیدا می‌شوند. بعد از آن، با استفاده از اطلاعات آماری به دست آمده از مجموعه برجسب‌های این نمونه‌های همسایه، یعنی تعداد نمونه‌های همسایه متعلق به هر کلاس ممکن، با استفاده از قانون حداکثر کردن احتمال پسین^{۲۵} مجموعه برجسب‌های نمونه دیده نشده تعیین می‌شوند.

جدول‌های ۲ تا ۸ به ترتیب مقایسه عملکرد الگوریتم‌های مورد آزمایش را از دید معیارهای کاهش ویژگی‌ها، دقت، افت همینگ، یک-خطا، پوشش، افت رتبه و میانگین صحت نشان می‌دهند. برای هر معیار ارزیابی، علامت "↑" به معنی "مقدار بیشتر نتیجه بهتر" و علامت "↓" به معنی "مقدار کمتر، نتیجه بهتر" است. در هر جدول، نتیجه بهتر به صورت پر رنگ‌تر نشان داده شده و زیر بدترین نتیجه خط کشیده شده است.

در جدول ۲ نتایج مقایسه الگوریتم‌های مختلف از دید معیار کاهش ویژگی‌ها آورده شده است. همانطور که گفته شد، این معیار مقادیری در بازه [۰، ۱] دارد و هر چه این مقدار به ۱ نزدیک‌تر باشد، میزان کاهش ویژگی‌ها بیشتر است. طبق این جدول، بازه تغییرات این معیار برای الگوریتم‌های مختلف بسیار متفاوت می‌باشد. به طور مثال، در مجموعه داده cal500 الگوریتم ABACO در حدود ۴۵٪ ویژگی‌ها را حذف کرده است؛ در حالیکه روش‌های فیلتری LP-RF و LP-IG نتوانسته‌اند هیچ یک از ویژگی‌ها را حذف کنند و معیار F_1 برای آن‌ها صفر به دست آمده است. همچنین، سه مورد از روش‌های فیلتری که در آن‌ها تمام ویژگی‌ها امتیازی کمتر از حد آستانه به دست آورده‌اند با علامت "-" نشان داده شده‌اند. توجه به این نکته ضروری است که حذف ویژگی‌ها در صورتی اهمیت دارد که کارایی طبقه‌بند افت نکند. در مورد مثال بالا، با وجود حذف ۴۵٪ از ویژگی‌ها با الگوریتم ABACO، نتایج به دست آمده برای دقت طبقه‌بند بهتر از نتایج به دست آمده توسط روش‌های فیلتری است، یعنی کسب دقت بالاتر با تعداد ویژگی‌های کمتر. در مقابل، در مجموعه داده emotions، روش فیلتری LP-IG حدود ۷۶٪ از ویژگی‌ها و الگوریتم تکاملی BACO حدود ۴۹٪ از ویژگی‌ها را حذف کرده است. اما دقت طبقه‌بندی که برای LP-IG و BACO به دست آمده است به ترتیب عبارتند از ۰/۵۹ و ۰/۶۷. بنابراین، یک روش انتخاب ویژگی خوب، روشی است که بتواند بین معیار کاهش ویژگی‌ها و کارایی طبقه‌بند یک مصالحه برقرار کند.

جدول ۲: مقایسه عملکرد الگوریتم‌های انتخاب ویژگی چند برچسبی بر حسب معیار کاهش ویژگی‌ها ↑

	ABACO	BACO	ACO	BGSA	catFishPSO	GA	BR_RF	BR_IG	LP_RF	LP_IG
cal500	۰/۴۵۰	۰/۴۴۷	۰/۸۰۱	۰/۴۹۱	۰/۴۸۶	۰/۴۷۰	۰/۱۰۰	-	۰	۰
emotions	۰/۴۸۵	۰/۴۹۶	۰/۷۳۷	۰/۴۹۹	۰/۴۷۴	۰/۵۰۵	۰/۳۰۸	۰/۲۱۱	۰/۱۷۶	۰/۷۶۸
medical	۰/۳۶۳	۰/۳۹۳	۰/۴۳۶	۰/۳۹۶	۰/۳۹۱	۰/۴۲۲	۰/۸۶۸	۰/۹۹۶	۰/۹۳۴	۰/۹۸۷
enron	۰/۵۰۱	۰/۵۰۳	۰/۵۲۶	۰/۵۰۴	۰/۴۹۷	۰/۴۹۵	۰/۰۲۶	۰/۹۹۷	۰/۰۰۳	-
yeast	۰/۳۸۳	۰/۴۱۷	۰/۴۲۲	۰/۴۰۷	۰/۳۹۵	۰/۴۳۲	۰/۴۵۰	۰/۸۹۷	۰/۰۶۹	-

جدول ۳: مقایسه عملکرد الگوریتم‌های انتخاب ویژگی چند برچسبی بر حسب معیار دقت ↑

	ABACO	BACO	ACO	BGSA	catFishPSO	GA	BR_RF	BR_IG	LP_RF	LP_IG
cal500	۰/۲۱۶	۰/۲۱۵	۰/۲۱۲	۰/۲۰۹	۰/۲۱۶	۰/۲۱۴	۰/۱۹۶	-	۰/۱۹۸	۰/۱۹۴
emotions	۰/۵۰۴۸	۰/۵۱۲۳	۰/۴۹۶۳	۰/۴۵۹۳	۰/۵۱۰۳	۰/۴۸۳۳	۰/۳۲۰۳	۰/۳۳۶۵	۰/۳۷۹۳	۰/۳۶۷۴
medical	۰/۶۷۴۲	۰/۶۷۳۳	۰/۵۹۶۵	۰/۶۳۴۹	۰/۶۸۱۹	۰/۶۳۸۸	۰/۵۵۴۹	۰/۴۲۵۷	۰/۵۸۶۳	۰/۵۹۰۰
enron	۰/۳۷۷۵	۰/۳۷۶۲	۰/۳۵۱۳	۰/۳۵۶۸	۰/۳۸۷۰	۰/۳۶۷۰	۰/۲۹۰۱	۰/۲۰۸۳	۰/۳۰۵۶	-
yeast	۰/۵۰۸۹	۰/۵۰۹۳	۰/۴۹۶۵	۰/۴۹۹۵	۰/۵۰۹۲	۰/۵۰۳۱	۰/۴۹۶۱	۰/۴۴۶۹	۰/۵۰۴۴	-

جدول ۴: مقایسه عملکرد الگوریتم‌های انتخاب ویژگی چند برچسبی بر حسب معیار افت همینگ ↓

	ABACO	BACO	ACO	BGSA	catFishPSO	GA	BR_RF	BR_IG	LP_RF	LP_IG
cal500	۰/۱۳۸	۰/۱۳۹	۰/۱۳۸	۰/۱۳۸	۰/۱۳۸	۰/۱۳۹	۰/۱۴۰	-	۰/۱۳۹	۰/۱۳۹
emotions	۰/۲۲۶۴	۰/۲۲۳۱	۰/۲۲۵۲	۰/۲۴۱۴	۰/۲۲۳۹	۰/۲۲۳۴	۰/۲۶۵۲	۰/۲۶۸۴	۰/۲۵۵۳	۰/۲۶۱۳
medical	۰/۰۱۳۶	۰/۰۱۳۰	۰/۰۱۴۸	۰/۰۱۳۷	۰/۰۱۳۱	۰/۰۱۳۹	۰/۰۱۵۹	۰/۰۱۸۸	۰/۰۱۵۰	۰/۰۱۴۹
enron	۰/۰۵۳۱	۰/۰۵۲۴	۰/۰۵۳۴	۰/۰۵۲۹	۰/۰۵۲۱	۰/۰۵۲۷	۰/۰۵۳۷	۰/۰۵۲۹	۰/۰۵۳۰	-
yeast	۰/۲۰۰۷	۰/۲۰۰۲	۰/۲۰۲۰	۰/۲۰۱۴	۰/۲۰۲۰	۰/۲۰۱۸	۰/۱۹۹۰	۰/۲۱۱۰	۰/۱۹۸۱	-

جدول ۵: مقایسه عملکرد الگوریتم‌های انتخاب ویژگی چند برچسبی بر حسب معیار یک-خطا ↓

	ABACO	BACO	ACO	BGSA	catFishPSO	GA	BR_RF	BR_IG	LP_RF	LP_IG
cal500	۰/۱۱۹	۰/۱۱۹	۰/۱۱۸	۰/۱۱۷	۰/۱۲۳	۰/۱۱۷	۰/۱۲۰	-	۰/۱۲۳	۰/۱۱۹
emotions	۰/۳۳۱۲	۰/۳۲۶۸	۰/۳۳۴۶	۰/۳۵۹۷	۰/۳۳۷۱	۰/۳۵۰۶	۰/۳۸۱۹	۰/۳۹۵۶	۰/۳۸۳۱	۰/۳۹۶۲
medical	۰/۲۲۸۱	۰/۲۲۸۰	۰/۲۶۴۵	۰/۲۴۸۷	۰/۲۲۶۱	۰/۲۴۸۱	۰/۲۶۷۳	۰/۴۲۸۱	۰/۲۴۴۵	۰/۲۷۰۳
enron	۰/۳۰۵۶	۰/۲۹۶۵	۰/۳۱۹۳	۰/۳۱۰۲	۰/۳۰۱۲	۰/۳۰۷۱	۰/۳۲۵۷	۰/۳۹۹۹	۰/۳۱۸۹	-
yeast	۰/۲۴۱۳	۰/۲۴۱۷	۰/۲۴۵۶	۰/۲۴۲۰	۰/۲۴۱۱	۰/۲۴۵۸	۰/۲۳۸۸	۰/۲۴۹۸	۰/۲۴۰۶	-

جدول ۶: مقایسه عملکرد الگوریتم‌های انتخاب ویژگی چند برچسبی بر حسب معیار پوشش ↓

	ABACO	BACO	ACO	BGSA	catFishPSO	GA	BR_RF	BR_IG	LP_RF	LP_IG
cal500	۱۳۱/۰۳۳	۱۳۰/۹۷۲	۱۳۰/۷۹۷	۱۳۱/۳۳۴	۱۳۱/۳۷۵	۱۳۰/۸۳۴	۱۳۱/۱۳۹	-	۱۳۰/۷۸۷	۱۳۰/۹۱۴
emotions	۲/۰۰۱۷	۱/۹۹۳۳	۱/۹۹۴۹	۲/۱۰۳۴	۱/۹۹۴۹	۲/۰۴۴۷	۲/۲۹۴۹	۲/۲۸۴۴	۲/۱۹۱۶	۲/۱۶۰۳
medical	۳/۰۲۴۲	۲/۹۱۸۵	۲/۹۴۹۶	۲/۸۹۷۳	۲/۷۹۶۹	۲/۸۵۰۸	۳/۱۲۲۹	۴/۸۱۶۲	۳/۰۵۹۶	۳/۱۳۳۱
enron	۱۳/۶۹۳۸	۱۳/۵۷۵۰	۱۳/۷۵۹۰	۱۳/۶۴۱۳	۱۳/۶۹۴۱	۱۳/۴۸۸۶	۱۳/۷۶۹۵	۱۴/۶۵۹۸	۱۳/۴۴۰۷	-
yeast	۶/۴۳۳۷	۶/۴۵۰۸	۶/۴۸۲۰	۶/۴۶۱۶	۶/۴۳۷۰	۶/۴۵۰۶	۶/۳۵۷۵	۶/۵۹۶۴	۶/۳۶۳۷	-

جدول ۷: مقایسه عملکرد الگوریتم‌های انتخاب ویژگی چند برجسیبی بر حسب معیار افت رتبه ↓

	ABACO	BACO	ACO	BGSA	catFishPSO	GA	BR_RF	BR_IG	LP_RF	LP_IG
cal500	۰/۱۸۶	۰/۱۸۶	۰/۱۸۶	۰/۱۸۶	۰/۱۸۶	۰/۱۸۷	۰/۱۸۵	-	۰/۱۸۵	۰/۱۸۲
emotions	۰/۲۰۷۸	۰/۲۰۰۳	۰/۲۰۴۹	۰/۲۲۶۱	۰/۲۰۸۰	۰/۲۱۷۲	۰/۲۶۲۳	۰/۲۶۲۰	۰/۲۴۶۷	۰/۲۳۸۸
medical	۰/۰۴۸۵	۰/۰۴۵۸	۰/۰۴۸۶	۰/۰۴۵۶	۰/۰۴۴۴	۰/۰۴۵۳	۰/۰۴۹۹	۰/۰۸۵۰	۰/۰۴۸۳	۰/۰۵۱۲
enron	۰/۰۹۵۵	۰/۰۹۵۰	۰/۰۹۷۶	۰/۰۹۶۲	۰/۰۹۶۴	۰/۰۹۵۰	۰/۰۹۷۳	۰/۱۰۹۱	۰/۰۹۵۷	-
yeast	۰/۱۷۷۷	۰/۱۷۷۴	۰/۱۸۰۲	۰/۱۷۸۹	۰/۱۷۷۶	۰/۱۷۹۴	۰/۱۷۳۰	۰/۱۸۷۶	۰/۱۷۳۷	-

جدول ۸: مقایسه عملکرد الگوریتم‌های انتخاب ویژگی چند برجسیبی بر حسب معیار میانگین صحت ↑

	ABACO	BACO	ACO	BGSA	catFishPSO	GA	BR_RF	BR_IG	LP_RF	LP_IG
cal500	۰/۴۸۶	۰/۴۸۸	۰/۴۸۴	۰/۴۸۶	۰/۴۸۷	۰/۴۸۶	۰/۴۸۸	-	۰/۴۸۸	۰/۴۸۸
emotions	۰/۷۵۹۷	۰/۷۷۳۷	۰/۷۵۷۵	۰/۷۳۹۱	۰/۷۵۶۲	۰/۷۴۸۳	۰/۷۱۰۷	۰/۷۰۷۵	۰/۷۱۹۷	۰/۷۱۹۷
medical	۰/۸۱۷۰	۰/۸۱۹۳	۰/۷۹۰۲	۰/۸۰۳۳	۰/۸۲۰۶	۰/۸۰۸۴	۰/۷۸۷۶	۰/۶۴۸۹	۰/۸۰۳۵	۰/۷۸۵۳
enron	۰/۶۳۲۳	۰/۶۳۴۷	۰/۶۲۱۷	۰/۶۲۷۸	۰/۶۳۱۸	۰/۶۳۰۱	۰/۶۱۴۹	۰/۵۴۴۴	۰/۶۱۹۹	-
yeast	۰/۷۵۰۳	۰/۷۵۱۳	۰/۷۴۶۷	۰/۷۴۸۶	۰/۷۵۰۵	۰/۷۴۷۱	۰/۷۵۵۹	۰/۷۳۵۷	۰/۷۵۵۱	-

جدول ۹: مقایسه عملکرد الگوریتم‌های انتخاب ویژگی چند برجسیبی بر حسب زمان پردازش ↓

	ABACO	BACO	ACO	BGSA	catFishPSO	GA	BR_RF	BR_IG	LP_RF	LP_IG
cal500	۲۸۳/۲۵۳	۳۱۰/۷۵۰	۱۸۶/۵۸۵	۲۰۲/۵۱۷	۲۰۸/۴۵۹	۲۱۷/۰۶۷	۷/۶۵۹	۳/۱۲۹	۰/۳۲۷	۰/۷۴۷
emotions	۱۵۶/۵۲	۱۴۲/۲۲۷	۱۱۶/۷۵۷	۱۱۷/۱۹۹	۱۸۳/۹۳۲	۹۸/۳۴۹	۰/۴۳۸	۰/۲۰۰	۰/۱۵۶	۰/۱۱۷
medical	۳۱۳/۴۸۳	۳۴۷/۸۶۹	۱۱۲/۲۷۶	۲۵۴/۳۸۰	۳۳۱/۰۱۴	۲۶۳/۷۴۰	۱۴۹/۲۶۱	۱۰/۴۸۰	۴/۸۳۹	۰/۳۹۸
enron	۳۷۳/۲۹۸	۲۵۴/۴۳۳	۱۳۹/۹۸۶	۲۱۷/۲۴۶	۳۱۴/۸۲۲	۱۹۶/۷۹۱	۲۸۱/۱۹۳	۱۵/۳۹۱	۱۱/۰۵۷	-
yeast	۸۱۴/۱۱۴	۹۱۱/۲۴۴	۴۵۲/۸۷۶	۸۷۳/۳۷۷	۹۷۵/۳۰۷	۷۰۴/۱۷۶	۱۴/۷۶۶	۲/۴۲۳	۲/۸۲۵	-

[۸] فاطمه علیقارداشی و محمدعلی زارع چاهوکی، "تأثیر ترکیب

روش‌های انتخاب ویژگی فیلتر و بسته‌بندی در بهبود پیش‌بینی اشکال نرم‌افزار، "مجله مهندسی برق دانشگاه تبریز، دوره ۴۷، شماره ۱، صفحات ۱۸۳ تا ۱۹۵، بهار ۱۳۹۶.

[۹] شیما کاشف و حسین نظام‌آبادی‌پور، "ارائه یک نسخه جدید از الگوریتم مورچگان باینری به منظور حل مسأله انتخاب ویژگی،" نشریه مهندسی برق و کامپیوتر ایران، دوره ۱۲، شماره ۲، صفحات ۱۲۷ تا ۱۴۴، زمستان ۱۳۹۳.

[10] S. Kashef and H. Nezamabadi-pour, "An advanced ACO algorithm for feature subset selection," *Neurocomputing*, vol. 147, pp. 271-279, 2015.

[۱۱] حامد توحیدی، حسین نظام‌آبادی‌پور و س. سریزدی، "انتخاب ویژگی با استفاده از الگوریتم جمعیت مورچگان باینری،" اولین کنگره مشترک سیستم‌های فازی و هوشمند، مشهد، ایران، ۱۳۸۶.

[12] E. Rashedi, H. Nezamabadi-Pour, and S. Saryazdi, "BGSA: binary gravitational search algorithm," *Natural Computing*, vol. 9, pp. 727-745, 2010.

[13] 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, pp. 12699-12707, 2011.

[14] N. Spolaôr, E. A. Cherman, M. C. Monard, and H. D. Lee, "A comparison of multi-label feature selection

مراجع

- [1] Q. Luo, E. Chen, and H. Xiong, "A semantic term weighting scheme for text categorization," *Expert Systems with Applications*, vol. 38, pp. 12708-12716, 2011.
- [2] K. Trohidis, G. Tsoumakas, G. Kalliris, and I. P. Vlahavas, "Multi-Label Classification of Music into Emotions," in *ISMIR*, pp. 325-330, 2008.
- [3] J. Yang, Y.-G. Jiang, A. G. Hauptmann, and C.-W. Ngo, "Evaluating bag-of-visual-words representations in scene classification," in *Proceedings of the international workshop on Workshop on multimedia information retrieval*, pp. 197-206, 2007.
- [4] M. R. Boutell, J. Luo, X. Shen, and C. M. Brown, "Learning multi-label scene classification," *Pattern recognition*, vol. 37, pp. 1757-1771, 2004.
- [5] S. Diplaris, G. Tsoumakas, P. A. Mitkas, and I. Vlahavas, "Protein classification with multiple algorithms," in *Panhellenic Conference on Informatics*, pp. 448-456, 2005.
- [6] M.-L. Zhang and Z.-H. Zhou, "Multilabel neural networks with applications to functional genomics and text categorization," *IEEE transactions on Knowledge and Data Engineering*, vol. 18, pp. 1338-1351, 2006.
- [7] S. Kashef and H. Nezamabadi-pour, "A new feature selection algorithm based on binary ant colony optimization," in *Information and Knowledge Technology (IKT), 2013 5th Conference on*, pp. 50-54, 2013.

- [26] J. Lee and D.-W. Kim, "Memetic feature selection algorithm for multi-label classification," *Information Sciences*, vol. 293, pp. 80-96, 2015.
- [27] Y. Lin, Q. Hu, J. Liu, and J. Duan, "Multi-label feature selection based on max-dependency and min-redundancy," *Neurocomputing*, vol. 168, pp. 92-103, 2015.
- [28] 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, pp. 1226-1238, 2005.
- [29] J. Yin, T. Tao, and J. Xu, "A Multi-label feature selection algorithm based on multi-objective optimization," in *Neural Networks (IJCNN), 2015 International Joint Conference on*, pp. 1-7, 2015.
- [30] N. Spolaôr, M. C. Monard, G. Tsoumakas, and H. D. Lee, "A systematic review of multi-label feature selection and a new method based on label construction," *Neurocomputing*, vol. 180, pp. 3-15, 2016.
- [31] H. Lim, J. Lee, and D.-W. Kim, "Optimization approach for feature selection in multi-label classification," *Pattern Recognition Letters*, vol. 89, pp. 25-30, 2017.
- [32] J. Lee and D.-W. Kim, "SCLS: Multi-label feature selection based on scalable criterion for large label set," *Pattern Recognition*, 2017.
- [33] L. Qiao, L. Zhang, Z. Sun, and X. Liu, "Selecting label-dependent features for multi-label classification," *Neurocomputing*, 2017.
- [34] L. Yu and H. Liu, "Feature selection for high-dimensional data: A fast correlation-based filter solution," in *ICML*, pp. 856-863, 2003.
- [35] J. Biesiada and W. Duch, "Feature selection for high-dimensional data—a Pearson redundancy based filter," in *Computer Recognition Systems 2*, ed: Springer, pp. 242-249, 2007.
- [36] C. G. Weng and J. Poon, "A new evaluation measure for imbalanced datasets," in *Proceedings of the 7th Australasian Data Mining Conference-Volume 87*, pp. 27-32, 2008.
- [15] M.-L. Zhang, J. M. Peña, and V. Robles, "Feature selection for multi-label naive Bayes classification," *Information Sciences*, vol. 179, pp. 3218-3229, 2009.
- [16] M.-L. Zhang and Z.-H. Zhou, "ML-KNN: A lazy learning approach to multi-label learning," *Pattern recognition*, vol. 40, pp. 2038-2048, 2007.
- [17] F. De Comité, R. Gilleron, and M. Tommasi, "Learning multi-label alternating decision trees from texts and data," in *International Workshop on Machine Learning and Data Mining in Pattern Recognition*, pp. 35-49, 2003.
- [18] E. Spyromitros, G. Tsoumakas, and I. Vlahavas, "An empirical study of lazy multilabel classification algorithms," in *Hellenic conference on Artificial Intelligence*, pp. 401-406, 2008.
- [19] L. Zhang, Q. Hu, J. Duan, and X. Wang, "Multi-label feature selection with fuzzy rough sets," in *International Conference on Rough Sets and Knowledge Technology*, pp. 121-128, 2014.
- [20] G. Doquire and M. Verleysen, "Feature selection for multi-label classification problems," in *International Work-Conference on Artificial Neural Networks*, pp. 9-16, 2011.
- [21] J. Read, B. Pfahringer, and G. Holmes, "Multi-label classification using ensembles of pruned sets," in *2008 Eighth IEEE International Conference on Data Mining*, pp. 995-1000, 2008.
- [22] O. Reyes, C. Morell, and S. Ventura, "Scalable extensions of the ReliefF algorithm for weighting and selecting features on the multi-label learning context," *Neurocomputing*, vol. 161, pp. 168-182, 2015.
- [23] J. Lee and D.-W. Kim, "Feature selection for multi-label classification using multivariate mutual information," *Pattern Recognition Letters*, vol. 34, pp. 349-357, 2013.
- [24] O. Reyes, C. Morell, and S. Ventura, "ReliefF-ML: an extension of reliefF algorithm to multi-label learning," in *Iberoamerican Congress on Pattern Recognition*, pp. 528-535, 2013.
- [25] N. Spolaôr, E. A. Cherman, M. C. Monard, and H. D. Lee, "relief for multi-label feature selection," *IEEE Brazilian Conference on Intelligent Systems (BRACIS)*, pp. 6-11, 2013.

زیر نویس ها

¹⁵ Multivariate mutual information

¹⁶ Incremental

¹⁷ Max-dependency and min-redundancy

¹⁸ Pearson product-moment correlation coefficient

¹⁹ Precision

²⁰ Recall

²¹ Mulan

²² Label cardinality (LC)

²³ Label density (LD)

²⁴ Accuracy

²⁵ Hamming loss

²⁶ One-error

²⁷ Coverage

²⁸ Ranking loss

¹ Multi-label learning

² Semantic annotation of images

³ Filter

⁴ Wrapper

⁵ Hybrid

⁶ Mutual information

⁷ Information gain

⁸ Correlation

⁹ Problem transformation

¹⁰ Algorithm adaptation

¹¹ Multi-label Naïve Bayes

¹² Unbalanced

¹³ Pruned problem transformation

¹⁴ Greedy

²⁹ Average precision

³⁰ Feature reduction

³¹ Symmetric difference

³² Feature reduction

³³ Crossover

³⁴ Mutation

³⁵ Maximum a posteriori (MAP)

³⁶ Off-line