

یک الگوریتم انتخاب ویژگی چند برجسیبی برخط با استفاده از اطلاعات متقابل چندگانه

مریم رحمانی‌نیا^(۱) سندس بهادری^(۲)

(۱) گروه مهندسی کامپیوتر، مرکز قصر شیرین، دانشگاه آزاد اسلامی، کرمانشاه، ایران*

(۲) گروه مهندسی کامپیوتر، واحد ایلام، دانشگاه آزاد اسلامی، ایلام، ایران

(تاریخ دریافت: ۱۴۰۲/۱۰/۲۵ تاریخ پذیرش: ۱۴۰۳/۰۴/۰۹)

چکیده

در حال حاضر، با حجم بزرگی از داده‌های آموزشی چندبرجسیبی که در آن هر نمونه آموزشی، دارای بیش از یک برجسب است، روبه‌رو هستیم. یکی از مشکلات اصلی این نوع داده‌های آموزشی، افزایش ناگهانی تعداد ویژگی‌ها است؛ که کارایی الگوریتم‌های داده‌کاوی را با مشکل مواجه کرده است. روش‌های انتخاب ویژگی چندبرجسیبی، با انتخاب یک زیرمجموعه مهم از ویژگی‌های اولیه، ابعاد این داده‌ها را کاهش می‌دهند. از سوی دیگر، در دنیای امروز، داده‌ها به صورت جریانی و به تدریج با گذر زمان، به مجموعه اضافه می‌شوند. بنابراین، الگوریتم‌های انتخاب ویژگی چندبرجسیبی باید بتوانند به صورت برخط استفاده شوند. اگرچه اخیراً الگوریتم‌های انتخاب ویژگی چندبرجسیبی برخط متعددی ارائه شده است، اما هیچ‌کدام به وابستگی بین برجسب‌ها جهت تعیین ارتباط ویژگی‌ها و برجسب‌ها در جریان داده‌ها توجه نکرده‌اند. به همین منظور، در این مقاله، یک روش انتخاب ویژگی چندبرجسیبی برخط ارائه شده است که با استفاده از اطلاعات متقابل چندگانه و محاسبه وابستگی میان برجسب‌ها، به فرآیند انتخاب ویژگی‌ها می‌پردازد. همچنین، برای نمایش کارایی روش پیشنهادی، چند آزمایش بر روی چندین الگوریتم انتخاب ویژگی و چند داده آموزشی چندبرجسیبی انجام شده است. نتایج به‌دست‌آمده نشان می‌دهد به‌طور متوسط الگوریتم پیشنهادی به ازای معیارهای افت همینگ، افت رتبه، پوشش، متوسط دقت و یک خطا به ترتیب ۱۵٪، ۶٪، ۱۹٪، ۷٪ و ۱۳٪ عملکرد بهتری نسبت به سایر الگوریتم‌های انتخاب ویژگی چندبرجسیبی داشته است.

کلمات کلیدی: روش‌های انتخاب ویژگی چند برجسیبی، داده‌های چند برجسیبی با جریان ویژگی، اطلاعات متقابل چندگانه

*عهده‌دار مکاتبات:

مریم رحمانی‌نیا

نشانی: گروه مهندسی کامپیوتر، مرکز قصر شیرین، دانشگاه آزاد اسلامی، کرمانشاه، ایران

پست الکترونیکی: ma.rahmaninia@gmail.com

۱-۱- بیان مسئله

اخیرا داده‌های آموزشی چندبرچسبی در بسیاری از برنامه‌های کاربردی دنیای واقعی (مانند دسته‌بندی اسناد متنی [۱])، دسته‌بندی تصاویر [۲]، دسته‌بندی ژن‌ها [۱] و تشخیص احساسات در موسیقی [۲]) مورد استفاده قرار گرفته است. همین امر باعث شده تا الگوریتم‌های یادگیری ماشین، در داده‌های آموزشی چندبرچسبی، مورد توجه بسیاری از محققان قرار بگیرند. در داده‌های آموزشی چندبرچسبی، هر نمونه آموزشی به‌طور همزمان با بیش از یک برچسب مرتبط است [۲]. به عنوان مثال، یک سند می‌تواند به چندین موضوع مختلف متعلق باشد، یا یک ژن می‌تواند با مجموعه‌ای از اهداف کاربردی مختلف مرتبط باشد و یا یک قطعه موسیقی را می‌توان به ژانرهای مختلفی اختصاص داد. چندین چالش اساسی درباره داده‌های آموزشی چندبرچسبی وجود دارد که در ادامه به آن‌ها اشاره می‌کنم.

۱. چالش اول داده‌های آموزشی چندبرچسبی، به خصوص زمانی که با داده‌های متنی و روزنامه‌ها سروکار داریم، تعداد زیاد ویژگی‌ها است. این تعداد می‌تواند به ده‌ها هزار ویژگی یا حتی بیشتر هم برسد، که برای جامعه پژوهشی چالش‌برانگیز است. تعداد زیاد ویژگی‌ها باعث مسئله ازدحام ابعاد می‌شود که هزینه محاسباتی و دقت الگوریتم‌ها و روش‌های یادگیری ماشین را تحت تأثیر قرار می‌دهد. همچنین، بسیاری از این ویژگی‌ها، ویژگی‌های نامربوط و اضافی هستند که به کارایی الگوریتم‌های یادگیری ماشین ضربه می‌زند.

۲. دومین چالش این است که در مقابل داده‌های آموزشی تک‌برچسبی که برچسب‌های آن‌ها دو به دو مستقل هستند، در داده‌های آموزشی چندبرچسبی، برچسب‌ها به یکدیگر وابسته و مرتبط هستند و پیش‌بینی درست تمامی برچسب‌های مرتبط با یک نمونه آموزشی سخت است. از طرف دیگر، وابستگی بین برچسب‌ها باعث می‌شود برچسب‌های ناشناخته‌ای که متعلق به برخی نمونه‌ها هستند نیز شناسایی شوند.

۳. چالش سوم مربوط به برخط و جریانی بودن داده‌های آموزشی می‌شود. همانطور که می‌دانیم در دنیای امروزه داده‌ها به صورت تدریجی و با گذر زمان به مجموعه داده‌ها اضافه می‌شوند و دسترسی به مجموعه داده‌ها در شروع فرایند انتخاب ویژگی معمولا غیرممکن است؛ پس الگوریتم‌های انتخاب ویژگی باید بدون هیچ دانش پیش زمینه‌ای به فرایند انتخاب ویژگی بپردازند.

۱-۲- ضرورت انجام تحقیق

تا به امروز، تعداد زیادی روش انتخاب ویژگی چندبرچسبی به منظور رفع این چالش ها ارائه شده است [۳-۵]. اگرچه بسیاری از این روش ها به صورت جزئی برخی از چالش ها را حل کرده اند، اما هنوز هیچ روشی به طور کامل تمامی مسائل مطرح شده را حل نکرده است. در [۳]، چندین روش انتخاب ویژگی چندبرچسبی ارائه شده است؛ که برخی از آن ها نیازمند محاسبات پیچیده برای استنتاج این وابستگی ها هستند و برخی دیگر این وابستگی ها را نادیده می گیرند. در [۵] یک روش انتخاب ویژگی چندبرچسبی با استفاده از الگوریتم بهینه سازی ازدحام ذرات و نظریه اطلاعات متقابل ارائه شده است. همچنین در [۶] یک الگوریتم انتخاب ویژگی چندبرچسبی تکاملی ارائه شده است که با استفاده از چندین جمعیت، به جلوگیری از محدود کردن تعامل هیبریدیزاسیون کمک می کند و زیرمجموعه نهایی از ویژگی را جست و جو می کند. در [۷]، یک روش انتخاب ویژگی چندبرچسبی ارائه شده است که با استفاده از یک دسته بند، به ارزیابی زیرمجموعه های انتخاب شده از ویژگی ها می پردازد. اگرچه روش های موجود به خوبی به مسئله انتخاب ویژگی در داده های آموزشی چندبرچسبی پرداخته اند، اما دارای یک سری معایب نیز هستند. به عنوان مثال، در همه این الگوریتم ها، نیاز است که قبل از شروع الگوریتم، تمام داده های آموزشی در دسترس باشند ولی در بسیاری از داده های آموزشی در دنیای واقعی، داده های آموزشی به صورت آنلاین و به تدریج به مجموعه داده اضافه می شوند، به طوری که انتظار داشتن داده های کامل برای آموزش عملاً غیرممکن است [۲]. به عنوان مثال در تجزیه و تحلیل تصاویر [۶]، توصیفگرهای متعددی به صورت پویا و با گذر زمان، برای توصیف جنبه های مختلف تصاویر مانند نمودار گرادیان درونی، نمودار رنگی و انتقال مقیاس ویژگی های غیرقابل تغییر، اضافه می شوند. همچنین، کشف حفره های موجود در تصاویر ماهواره ای از مریخ، مثال دیگری از دنیای واقعی است که در آن ویژگی ها به تدریج به داده های آموزشی اضافه می شوند [۷].

همچنین روش های انتخاب ویژگی مختلفی برای انتخاب ویژگی ها به صورت برخط ارائه شده است [۹،۸]. به عنوان مثال، یکی از این روش ها در [۱۰] براساس نظریه اطلاعات ارائه شده است که به صورت توزیع شده، ویژگی های مهم در داده های چندبرچسبی را انتخاب می کند. در [۱۱] نیز یک روش جدید بر اساس اطلاعات متقابل فازی، به منظور انتخاب ویژگی ها در این داده ها معرفی شده است که به فرایند انتخاب ویژگی به صورت برخط می پردازد. همچنین، [۱۲] یک روش انتخاب ویژگی چندبرچسبی برخط را با استفاده از نظریه مجموعه های سخت ارائه می دهد. اگر چه این دسته از روش های انتخاب ویژگی برخط تا حدی چالش جریانی بودن داده های آموزشی را برطرف کرده است اما در هیچ کدام وابستگی میان برچسب ها در نظر گرفته نشده است. به منظور حل این مشکلات، در این مقاله، یک روش جدید به نام

MLOSFS^۱ به منظور انتخاب ویژگی‌های مهم در جریان داده‌های چندبرچسبی با استفاده از نظریه اطلاعات متقابل چندگانه ارائه شده است. در روش پیشنهادی، از مفهوم نظریه اطلاعات متقابل چندگانه استفاده می‌شود و تعداد ویژگی‌ها به صورت پیش‌فرض تعیین نمی‌شود؛ بلکه توسط الگوریتم به صورت خودکار تعیین می‌شود. این روش دارای دو فاز است: فاز ارتباط و فاز افزونگی. در فاز ارتباط، میزان وابستگی و ارتباط ویژگی‌های جدید با مجموعه برچسب‌ها محاسبه می‌شود. اگر ویژگی جدید وارد با مجموعه برچسب‌ها وابستگی و ارتباط داشته باشد، آنگاه ویژگی جدید به مجموعه ویژگی‌های انتخاب شده، افزوده می‌شود و الگوریتم وارد فاز افزونگی می‌شود. در غیر این صورت، الگوریتم منتظر ورود ویژگی جدید می‌ماند. در فاز افزونگی، وابستگی میان مجموعه ویژگی‌ها محاسبه می‌شود و ویژگی‌هایی که دارای افزونگی هستند (یعنی هیچ اطلاعات جدیدی در مورد مجموعه برچسب‌ها تولید نمی‌کنند)، حذف می‌شوند. در محاسبه میزان افزونگی، از وابستگی میان مجموعه برچسب‌ها و مجموعه ویژگی‌های قبلاً انتخاب شده با استفاده از اطلاعات متقابل چندمتغیره و اطلاعات متقابل شرطی استفاده می‌شود. ساختار این مقاله به این صورت است که در ابتدا یک پیش‌زمینه در مورد نظریه اطلاعات متقابل و پیشینه تحقیق در بخش ۲ ارائه می‌شود. در بخش ۳، الگوریتم پیشنهادی معرفی می‌شود. در بخش ۴، نتایج آزمایشگاهی به همراه تجزیه و تحلیل نتایج آورده شده است. در نهایت نتیجه‌گیری و کارهای آینده در فصل ۵ ارائه می‌شود.

۲- پیشینه تحقیق و پیش‌زمینه

۲-۱- پیشینه تحقیق: مروری بر کارهای انجام‌شده

به طور کلی، دو نوع از الگوریتم‌ها برای کاهش ابعاد داده‌های آموزشی حجیم چندبرچسبی وجود دارند. این الگوریتم‌ها را می‌توان در دو دسته استخراج ویژگی و انتخاب ویژگی قرار داد. روش‌های استخراج ویژگی چندبرچسبی، با استفاده از روش‌های نگاشت فضای ویژگی یا روش‌های انتقال فضا، ابعاد داده‌های آموزشی را کاهش می‌دهند [۱۳، ۱۴]. در این دسته از روش‌ها، معنای اولیه ویژگی‌ها و برچسب‌ها از بین می‌رود. به همین دلیل است که در محیط‌هایی که معنای اصلی و اولیه ویژگی‌ها لازم است، به خوبی عمل نمی‌کنند. در مقابل، در روش‌های انتخاب ویژگی چندبرچسبی، الگوریتم‌ها تنها سعی می‌کنند با انتخاب زیرمجموعه‌ای مهم و کارا از میان مجموعه ویژگی‌های اولیه، اندازه داده‌های آموزشی را کاهش دهند. در این روش‌ها، معنا و شکل اولیه ویژگی‌ها حفظ می‌شود و هیچ‌گونه تغییری در ویژگی‌ها به وجود نمی‌آید و تنها

^۱ Mutual Information-based Online Feature Selection

تعدادی از ویژگی‌های اضافی و نامربوط حذف می‌شوند. در ادامه، تعدادی از روش‌های استخراج ویژگی و انتخاب ویژگی را بیان می‌کنیم.

• روش‌های استخراج ویژگی چندبرچسبی

یکی از روش‌های معروف کاهش بعد بر مبنای استخراج ویژگی، روش تجزیه و تحلیل تبعیض‌آمیز خطی است [۱۳]. هدف این روش، یافتن یک تابع نگاشت خطی است که شباهت درون کلاسی را بیشینه کرده و در عین حال شباهت میان کلاس‌ها را کمینه کند. این روش اخیراً به چندین شکل در داده‌های آموزشی چندبرچسبی گسترش یافته است [۱۵، ۱۶]. تفاوت اصلی این روش‌ها به نحوه وزن‌دهی برچسب‌ها برمی‌گردد. به این ترتیب، روش [۴] از وزن‌دهی دودویی، روش [۳] از وزن‌دهی براساس آنتروپی، روش [۱۵] از وزن‌دهی براساس همبستگی، روش [۱۷] از وزن‌دهی براساس منطق فازی و روش [۱۶] از وزن‌دهی براساس وابستگی استفاده کرده است. در [۱۸]، یک روش استخراج ویژگی تحت نظارت، برای کاهش ابعاد داده‌های آموزشی مطرح شده است. این روش با استفاده از معیار استقلال هیلبرت-اشمیت [۱۹]، وابستگی بین ویژگی‌ها و برچسب‌ها را بیشینه می‌کند. این راهکار به دو صورت $MDDM_p$ و $MDDM_f$ معرفی شده است. روش $MDDM_p$ با استفاده از جهت‌های نگاشت متعامد عمل می‌کند، در حالی که روش $MDDM_f$ ویژگی‌های نگاشت‌شده را به صورت متعامد کاهش می‌دهد. در [۲۰]، نویسندگان نشان دادند که روش $MDDM_p$ را می‌توان به عنوان یک مسئله کمترین مربعات خطا فرموله کرد. آنها همچنین با ادغام روش تحلیل مولفه اصلی و تابع هدف روش $MDDM_p$ ، یک روش جدید به نام $MVMD$ را ارائه کردند. در [۲۱]، یک روش انتخاب ویژگی با استفاده از ابرگراف بر مبنای داده‌ها و برچسب‌ها معرفی شده است؛ که وابستگی بین برچسب‌های مختلف را بررسی می‌کند. در این روش، از فرمول یادگیری طیفی ابرگراف برای دسته‌بندی چندبرچسبی استفاده می‌شود. همچنین در [۲۲] با در نظر گرفتن ماتریس معکوس مور-پنروز^۱، هسته خطی^۲ برای فضای ویژگی و هسته دلتا^۳ برای فضای برچسب، مجموعه داده‌های آموزشی به مجموعه با اندازه کوچکتر نگاشت شده است.

• روش‌های انتخاب ویژگی چندبرچسبی

در مقابل، روش‌های انتخاب ویژگی چندبرچسبی ایستا، که در آنها اندازه داده‌های آموزشی ثابت است، تنها با انتخاب زیرمجموعه‌ای از ویژگی‌های موجود و با حفظ ساختار داده‌ها، کاهش ابعاد داده‌های آموزشی را انجام می‌دهند [۲۳-۲۷].

¹ Moore-Penrose

² Linear kernel

³ Delta kernel

اگرچه این روش‌ها به خوبی در داده‌های آموزشی چندبرچسبی مشکل ازدحام ابعاد را حل می‌کنند، اما وابستگی بین برچسب‌ها را در نظر نمی‌گیرند. از طرف دیگر، بیشتر این روش‌ها برای محاسبه بردارهای ویژه در ماتریس‌های بزرگ و چگال استفاده می‌کنند که محاسباتی بسیار زمان‌بر و هزینه‌بر در داده‌های با ابعاد بالا می‌باشد. به علاوه، می‌توان گفت که این روش‌ها یک مشکل مشترک دارند و آن این است که برای شروع الگوریتم، نیاز به دسترسی به کل مجموعه داده‌ها است. در بسیاری از برنامه‌های کاربردی واقعی، داده‌ها به طور پیوسته با گذر زمان به مجموعه اضافه می‌شوند و انتظار نمی‌رود که مجموعه داده‌ها به طور کامل در دسترس باشد. به عنوان مثال، در شبکه‌های اجتماعی، عناوین داغ و مورد توجه به صورت پیوسته اضافه می‌شوند و هنگام به‌روزرسانی عنوان جدید، کلمات کلیدی جدیدی نیز به داده‌ها اضافه می‌شود [۲۸].

برخلاف روش‌های انتخاب ویژگی ایستا، روش‌های انتخاب ویژگی برخط، ویژگی‌ها را در یک حالت برخط انتخاب می‌کنند [۳۰، ۲۹]. اما همه این روش‌ها فرض می‌کنند که هر نمونه تنها به یک برچسب تعلق دارد. همان‌طور که قبلاً گفته شد، در دنیای واقعی، هر نمونه می‌تواند به چندین برچسب تعلق داشته باشد. روش‌های آنلاین انتخاب ویژگی چندبرچسبی، فرض می‌کنند که با گذر زمان ویژگی‌ها به داده‌های آموزشی با بیش از یک برچسب اضافه می‌شوند. در [۱۲] یک روش چندبرچسبی آنلاین انتخاب ویژگی گروهی ارائه شده است. این روش شامل دو فاز است: فاز انتخاب گروهی و فاز انتخاب داخل گروهی. در فاز اول، گروه‌های موثر از میان گروه‌های اضافه شده تا آن لحظه، شناسایی می‌شوند و سپس در فاز دوم، ویژگی‌های مهم از هر گروه انتخاب شده و ویژگی‌های اضافی حذف می‌شوند. در این روش، میزان اهمیت یک گروه از ویژگی‌ها در فاز انتخاب گروهی و همچنین میزان ارتباط ویژگی‌ها در فاز انتخاب داخل گروهی محاسبه می‌شود. همچنین، در [۱۱]، دو روش برخط انتخاب ویژگی چندبرچسبی با استفاده از اطلاعات متقابل فازی ارائه شده است. در [۳۱]، نویسندگان یک روش آنلاین انتخاب ویژگی چندبرچسبی با استفاده از نظریه همسایگی مجموعه‌های سخت ارائه داده‌اند. این نظریه تنها برای داده‌های گسسته در داده‌های آموزشی استفاده می‌شود و استفاده از آن در داده‌های پیوسته با هزینه محاسباتی بالا همراه است. اخیراً، در [۲۶] یک روش برخط انتخاب ویژگی چندبرچسبی ارائه شده است که با استفاده از یک استراتژی جست‌وجوی چند هدفه و بهره‌گیری از نظریه مجموعه‌های پارتو و نظریه اطلاعات متقابل، به انتخاب ویژگی‌های موثر پرداخته می‌شود. در [۳۲] یک روش انتخاب ویژگی چندبرچسبی در جریان داده‌های برخط با استفاده از نظریه همسایگی مجموعه‌های سخت ارائه شده؛ که در آن اهمیت ویژگی، تکرار ویژگی و یکپارچگی فضای برچسب به طور همزمان مورد توجه قرار گرفته شده است. در [۳۳] نیز با استفاده از ارزیابی ارتباط قابل مقیاس، ارتباط مشروط ویژگی‌ها با مجموعه برچسب‌ها را به طور دقیق محاسبه می‌کند. در [۳۴]، یک روش سریع برای انتخاب ویژگی‌های چندبرچسبی بر اساس رتبه‌بندی ویژگی‌های اطلاعاتی معرفی شده است که از یک استراتژی

جدید به منظور انتخاب برچسب‌های مهم و در نتیجه کاهش میزان محاسبات استفاده کرده است. در [۳۵] نیز یک روش انتخاب ویژگی با استفاده از نظریه اطلاعات متقابل تقریبی بر مبنای نظریه نامساوی شیرر [۳۶] ارائه شده است که بدون انتقال مجموعه ویژگی‌ها و به صورت مستقیم ارتباط میان برچسب‌ها و ویژگی‌ها را محاسبه می‌کند. در [۳۷] یک روش انتخاب ویژگی چندبرچسبی دیگر با استفاده از نظریه اطلاعات متقابل شرطی ارائه شده است که تمرکز اصلی آن روی کاهش تعداد محاسبات اصلی است.

۲-۲- پیش‌زمینه: نظریه اطلاعات متقابل

تلاش‌های زیادی به منظور به دست آوردن مقدار اطلاعات متقابل [۳۸] میان مجموعه‌ای از متغیرها صورت گرفته است. این تلاش‌ها با نگرانی‌های زیادی روبه‌رو بوده است. از جمله آن که تعاملات میان بسیاری از متغیرها قابل درک نیست. از طرف دیگر، محاسبه همه تعاملات کار دشواری است. بر اساس [۳۸] اطلاعات متقابل میان زیرمجموعه‌ای از ویژگی‌ها مانند $S = \{f_1, f_2, \dots, f_n\}$ و مجموعه برچسب‌های $L = \{l_1, \dots, l_t\}$ عبارت است از:

$$\begin{aligned} I(S, L) &= H(S) - H(S, L) + H(L) \\ &= H(\{f_1, f_2, \dots, f_n\}) - H(\{f_1, f_2, \dots, f_n, l_1, \dots, l_t\}) + H(\{l_1, l_2, \dots, l_t\}) \end{aligned} \quad (1)$$

به طوری که هر کدام از معیارهای $H(\cdot)$ به آنتروپی مشترک میان مجموعه‌ای از متغیرها اشاره می‌کند که طبق رابطه زیر به دست می‌آید:

$$H(f_1, f_2, \dots, f_n) = - \sum P(\{f_1, f_2, \dots, f_n\}) \cdot \log P(\{f_1, f_2, \dots, f_n\}) \quad (2)$$

همچنین $P(\{f_1, f_2, \dots, f_n\})$ به تابع جرم احتمالی به ازای مجموعه متغیرهای $\{f_1, f_2, \dots, f_n\}$ اشاره می‌کند. آنتروپی معیاری است که مقدار اطلاعات درون یک متغیر را نشان می‌دهد. به منظور سادگی کار می‌توان گفت اطلاعات متقابل به میزان اطلاعات مشترک یک ویژگی با ویژگی‌های دیگر اشاره می‌کند. به طور کلی با استفاده از تعریف آورده شده در [۳۹] می‌توان اطلاعات متقابل چندگانه را به صورت حاصل جمع آنتروپی‌ها همانند زیر نوشت:

$$\begin{aligned} I(f_1; f_2 \dots f_n) &= \{H(f_1) + \dots + H(f_n)\} \\ &- \{H(f_1, f_2) + \dots + H(f_{n-1}, f_n)\} + \dots + (-1)^n H(f_1, \dots, f_n) \\ &= - \sum_{f \in T} (-1)^{|f|} H(T) \end{aligned} \quad (3)$$

به طوری که $\sum_{f \subseteq T'} f$ مجموع تمام عناصر مشتق شده از T' و T' به مجموعه توانی اعضای T اشاره می کند. به عنوان مثال فرض کنید $T = \{f_1, f_2\}$ ، پس مجموعه توانی آن عبارت است از $T' = \{\emptyset, f_1, f_2, \{f_1, f_2\}\}$ درحالی که اطلاعات متقابل دوتایی، وابستگی میان دو ویژگی را محاسبه می کند، اطلاعات متقابل چندگانه، میزان وابستگی میان چندین ویژگی را محاسبه می کند؛ اما چون محاسبه میزان آنتروپی مشترک مجموعه ای از ویژگی ها بسیار پرهزینه است، تلاش های زیادی به منظور ارائه یک تخمین مناسب از اطلاعات متقابل انجام شده است فرض کنید S به مجموعه ای از ویژگی ها با n ویژگی و f به یک عضو از مجموعه $S_k = \{e | e \subseteq S, |e| = k\}$ اشاره کند. حال مجموع آنتروپی همه زیرمجموعه های k عضوی از مجموعه S_k با استفاده از نماد U_k به صورت زیر نشان داده می شود:

$$U_k(S) = \sum_{f \in S_k} H(f) \quad (4)$$

از این نمادگذاری هنگامی که با مجموعه توانی سروکار داریم بسیار می توان استفاده کرد. حال با این فرض که $S = \{f_1, f_2, f_3\}$ و S' مجموعه توانی S باشد، حاصل جمع آنتروپی مجموعه های ۲ عضوی روی مجموعه S' عبارت است از:

$$U_2(S') = H(f_1; f_2) + H(f_1; f_3) - H(f_2; f_3) \quad (5)$$

حال اگر S'_k به مجموعه تمام زیرمجموعه های k عضوی از مجموعه توانی S' همانند زیر اشاره کند:

$$S'_k = \{e | e \in S', |e| = k\} \quad (6)$$

در این صورت، با در نظر گرفتن Y به عنوان یک عضو از مجموعه X'_m ، رابطه شماره (۲) را می توان به صورت زیر بازنویسی کرد:

$$H(S) = \sum_{k=1}^n \sum_{m=1}^k (-1)^{k+m} \left(\sum_{X \in S'_k} \sum_{Y \in X'_m} H(Y) \right) \quad (7)$$

حال به منظور تخمین آنتروپی توام چندین متغیر با استفاده از آنتروپی های k عضوی، می توان حاصل جمع بالا را به صورت زیر نوشت:

$$H(S) = - \sum_{X \in S'} (-1)^{|X|} \left(- \sum_{Y \in X'} (-1)^{|Y|} H(Y) \right) = - \sum_{X \in S'} (-1)^{|X|} I(\{X\}) \quad (8)$$

از طرف دیگر همانند فرمول رابطه (۴)، می توان مجموع اطلاعات متقابل با زیرمجموعه‌هایی با سایز k را به صورت زیر نوشت:

$$V_k(S) = \sum_{f \in S_k} I(\{X\}) \quad (9)$$

در نتیجه فرمول رابطه (۷) را به صورت زیر می توان بازنویسی کرد:

$$H(S) = - \sum_{k=1}^n (-1)^k V_k(S) \quad (10)$$

به طور کلی می توان نتیجه گرفت که آنتروپی یک مجموعه را، می توان با استفاده از فرمول بالا و روی مجموعه‌های k عضوی تعریف کرد. حال اگر L به مجموعه برچسب‌ها با حداکثر اندازه t اشاره کند، می توان آنتروپی L را به صورت زیر نوشت:

$$H(L) = - \sum_{k=1}^t (-1)^k V_k(L) \quad (11)$$

با استفاده از رابطه شماره (۱۱) می توان آنتروپی توام دو مجموعه S و L را به صورت آنتروپی توام روی مجموعه های k عضوی از مجموعه $\{S, L\}$ همانند زیر تعریف کرد:

$$H(S, L) = - \sum_{k=1}^{(n+t)} (-1)^k V_k(S, L) \quad (12)$$

به طوری که $\{S, L\}'$ به مجموعه توانی حاصل ضرب دکارتی مجموعه‌های S و L اشاره می کند. مجموعه توانی $\{S, L\}'$ در طرف راست رابطه (۱۲-۵) را می توان به صورت سه عبارت نوشت. به این صورت که عبارت اول از حاصل ضرب دکارتی $S' \times L'_0$ عبارت دوم از حاصل ضرب مجموعه‌های $S'_0 \times L'$ و عبارت سوم از حاصل ضرب اعضای باقی مانده مجموعه‌ها به دست می آید. همان طور که قبلاً گفته شد، L'_0 و S'_0 به عضوهای صفر عضوی از مجموعه توانی L' و S' اشاره می کند. به عنوان مثال، $V_3(\{S, L\}')$ را می توان به صورت حاصل جمع سه عبارت نوشت: عبارت

اول $V_3(S'_3 \times L'_0)$ ، عبارت دوم $V_3(S'_0 \times L'_3)$ و عبارت سوم $V_3(\{S'_2 \times L'_1\}) + V_3(\{S'_1 \times L'_2\})$ می‌توان گفت $V_k(\{S, L\}')$ عبارت است از:

$$V_k(\{S, L\}') = V_k(S'_k \times L'_0) + V_k(S'_0 \times L'_k) + \sum_{p=1}^{(k-1)} V_p(S'_{k-p} \times L'_p) \quad (13)$$

عبارت بالا نشان می‌دهد که هر زیرمجموعه k عضوی از مجموعه $\{S, L\}'$ را می‌توان به صورت اعضای

متغیرهای انتخاب شده از ترکیب S و L را شامل می‌شود، پس شامل اعضای با اندازه بیشتر یا مساوی دو است. در نتیجه با استفاده از رابطه (13) می‌توان رابطه (12) را به صورت زیر نوشت:

$$\begin{aligned} H(S, L) &= - \sum_{k=1}^{(n+t)} (-1)^k \left(V_k(S'_k \times L'_0) + V_k(S'_0 \times L'_k) \right) \\ &\quad - \sum_{k=2}^{(n+t)} (-1)^k \left(\sum_{p=1}^{(k-1)} V_p(S'_{k-p} \times L'_p) \right) \\ &= - \sum_{k=1}^n (-1)^k V_k(S') - \sum_{k=1}^t (-1)^k V_k(L') \\ &\quad - \sum_{k=2}^{(n+t)} \sum_{p=1}^{(k-1)} (-1)^k V_p(S'_{k-p} \times L'_p) \end{aligned} \quad (14)$$

به طوری که n و t به ترتیب به اندازه مجموعه L و S اشاره می‌کند. در نتیجه با استفاده از رابطه (14) اطلاعات متقابل

میان دو مجموعه را می‌توان به صورت زیر خلاصه کرد:

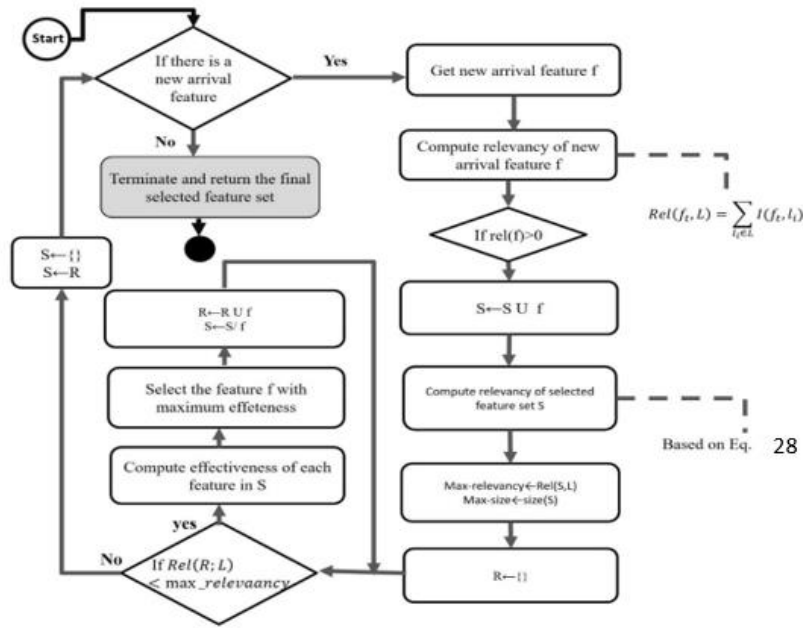
$$\begin{aligned} I(S, L) &= H(S) + H(L) - H(S, L) \\ &= - \sum_{k=1}^t (-1)^k V_k(S') - \sum_{k=1}^t (-1)^k V_k(L') + \sum_{k=1}^n (-1)^k V_k(S') + \sum_{k=1}^t (-1)^k V_k(L') \\ &\quad + \sum_{k=2}^{n+t} \sum_{p=1}^{k-1} (-1)^k V_k(S'_{k-p} \times L'_p) = \sum_{k=2}^{n+t} \sum_{p=1}^{k-1} (-1)^k V_k(S'_{k-p} \times L'_p) \end{aligned} \quad (15)$$

۳- راهکار پیشنهادی

برای انتخاب ویژگی در الگوریتم‌های انتخاب ویژگی چندبرچسبی، مجموعه داده‌های آموزشی را با نماد $X = R^{n \times d}$ نشان می‌دهیم. در اینجا، n تعداد نمونه‌های آموزشی و d تعداد ویژگی‌ها را نشان می‌دهد. همچنین، فرض کنید $L = \{-1, +1\}^{|L|}$ به مجموعه فضای برچسب‌ها با $|L|$ حالت مختلف اشاره کند. فضای ویژگی‌ها نیز یک بردار با جریان پویا $F = \{f_1, f_2, \dots, f_i, \dots\}$ است که به تدریج به مجموعه داده‌ها اضافه می‌شود. در این قسمت، یک راهکار برای انتخاب ویژگی مبتنی بر نظریه اطلاعات متقابل چندگانه در داده‌های آموزشی چندبرچسبی ارائه می‌دهیم. این راهکار قادر است با وجود جریان ویژگی، از میان ویژگی‌هایی که تا آن لحظه وارد شده‌اند، یک زیرمجموعه بهینه از ویژگی‌ها را انتخاب کند.

۳-۱- چهارچوب روش پیشنهادی

در این بخش، الگوریتم پیشنهادی به نام MLOSFS (Multi Label Online Stream Feature Selection) که برای انتخاب ویژگی در داده‌های چندبرچسبی با جریان آنلاین طراحی شده است را معرفی می‌کنیم. تعداد ویژگی‌ها در این الگوریتم به طور خودکار توسط خود الگوریتم محاسبه می‌شود. همان‌طور که در شکل ۱، چهارچوب روش پیشنهادی هنگامی که یک ویژگی وارد می‌شود، MLOSFS ابتدا ارتباط آن با برچسب‌ها را بررسی می‌کند. اگر ویژگی با برچسب‌ها مرتبط باشد، آن را به مجموعه ویژگی‌های انتخاب شده S اضافه می‌کند؛ در غیر این صورت، آن را نادیده می‌گیرد. اگر ویژگی به مجموعه S اضافه شود، MLOSFS تعدادی از ویژگی‌هایی که هیچ تأثیری در پیش‌بینی برچسب‌ها ندارند را حذف می‌کند. این فرایند تا زمانی ادامه می‌یابد که ویژگی جدیدی به داده‌ها اضافه نشود. به طور کلی، می‌توان گفت که این الگوریتم دارای دو فاز اصلی به نام ارتباط و افزونگی است. در شکل ۱، چهارچوب کلی روش پیشنهادی پیشنهادی آورده شده است.



شکل ۱: فلوجارت الگوریتم پیشنهادی

فاز اول: در مرحله اول، یعنی فاز ارتباط، ما بررسی می‌کنیم که آیا ویژگی جدید با برچسب‌های کلاس مرتبط است یا خیر؟ در یادگیری چندبرچسبی، میزان ارتباط یک ویژگی جدید f_i با برچسب‌های کلاس (L) با استفاده از محاسبه مقدار $I(f_i; L)$ به صورت زیر تعیین می‌شود.

$$I(f_i; L) = H(f_i) - H(f_i, L) + H(L) \quad (16)$$

به منظور جلوگیری از احتمال تخمین نادرست ناشی از مجموعه داده‌های چند برچسبی با ابعاد بزرگ، با در نظر گرفتن

$$|L| = t, \text{ رابطه بالا را می‌توان به صورت زیر تخمین زد:}$$

$$I(f_i; L) = \sum_{k=2}^{1+t} \sum_{p=1}^{k-1} (-1)^k V_k(f'_{i(k-p)} \times L'_p) = \sum_{k=2}^{1+t} (-1)^k V_k(f_i \times L'_{k-1}) \quad (17)$$

به طوری که \times به حاصل ضرب دکارتی و $V_k(\cdot)$ به مجموع تعاملات با درجه k میان متغیرها اشاره دارد که طبق رابطه

(۹) عبارت است از:

$$V_k(L) = \sum_{X \in L_k} I(X) \quad (18)$$

L' به مجموعه توانی از L ، بدون در نظر گرفتن $\{\emptyset\}$ ، L'_k به تمام زیرمجموعه‌های k عضوی از مجموعه توانی L' و X به یک عضو دلخواه از مجموعه توانی L'_k با اندازه k اشاره می‌کند. همچنین مقدار $I(X)$ عبارت است از:

$$I(X) = - \sum_{Y \in X} (-1)^{|Y|} H(Y) \quad (19)$$

به منظور محاسبه دقیق این عبارت تمام حالات ممکن از زیر مجموعه‌های مجموعه توانی L به منظور محاسبه اطلاعات متقابل میان ویژگی f_i و مجموعه برجسب‌های L باید در نظر گرفته شود. از آنجا که $t = |f_i| = 1$ پس می‌توان رابطه زیر را به دست آورد:

$$I(f_i; L) = \sum_{k=2}^{1+b} (-1)^k V_k(f_i \times L'_{k-1}) \quad (20)$$

به همین ترتیب یک تخمین ساده از رابطه بالا عبارت است از:

$$I(f_i; L) = \sum_{k=2}^2 (-1)^k V_k(f_i \times L'_{k-1}) = \sum_{l \in L} I(f_i; l_i) \quad (21)$$

طبق این الگوریتم اگر ویژگی تازه‌وارد با مجموعه برجسب‌ها بر اساس رابطه (۲۱) مرتبط بود، ویژگی تازه‌وارد به مجموعه ویژگی‌های انتخابی اضافه و الگوریتم وارد فاز دوم می‌شود (خط ۴ تا ۷ الگوریتم پیشنهادی در شکل ۲). در غیر این صورت الگوریتم منتظر ورود ویژگی جدید می‌ماند.

Algorithm 1: Multi Label Online Stream Feature Selection

Input:

L: the class labels
 f_t : the new arrival feature in time t

Output:

R: the set of final selected features

1. **Do**
2. **//phase 1: relevance**
3. Get new arrival feature f_t
4. $Rel(f_t, L) = \sum_{l_i \in L} I(f_t, l_i)$
5. **If** $Rel(f_t, L) > 0$
6. $S \leftarrow S \cup \{f_t\}$
7. **End if**
8. **//phase 2: redundancy**
9. $Rel(S, L) = \sum_{f_i \in S} \sum_{l_i \in L} I(f_i, l_i)$
10. $Max\text{-relevancy} \leftarrow Rel(S, L)$
11. $Max\text{-size} \leftarrow size(S)$
12. $R \leftarrow \{\}$
13. $Rel(R, L) = 0$
14. **While** $Rel(R, L) < Max\text{-relevancy}$ and $size(R) < Max\text{-size}$
15. For each feature f in S
16. Compute effectiveness f based on Eq. (28)
17. End for
18. A \leftarrow Select the feature with maximum effectiveness
19. $R \leftarrow R \cup A$
20. $S \leftarrow S - A$
21. **End while**
22. **Until** there is no new arrival feature
23. **Return** R

شکل ۲: شبه کد الگوریتم پیشنهادی

فاز دوم، افزونگی: در این فاز قدرت تاثیر ویژگی‌ها بر اساس خطوط ۸ تا ۲۱ الگوریتم پیشنهادی در شکل ۲ محاسبه می‌شود. هدف اصلی در این فاز، انتخاب حداقل تعداد ویژگی است به طوری که میزان ارتباط ویژگی‌ها همچنان حفظ شود. در این فاز در هر مرحله ویژگی‌هایی با قدرت تاثیر بالا انتخاب می‌شوند. میزان تاثیر یک ویژگی را می‌توان با در نظر گرفتن ارتباط آن ویژگی، با برچسب کلاس و در نظر گرفتن سایر ویژگی‌های انتخاب شده قبلی محاسبه کرد. هر چه میزان ارتباط یک ویژگی با مجموعه برچسب‌ها بیشتر و افزونگی کمتری داشته باشد، پس قدرت تاثیر ویژگی بیشتر است. انتخاب ویژگی با قدرت تاثیر بالا از میان مجموعه ویژگی‌های انتخاب شده تا زمانی ادامه می‌یابد، که میزان ارتباط مجموعه ویژگی‌های انتخاب شده جدید (R) با برچسب کلاس نزدیک به مقدار ارتباط مجموعه ویژگی‌های انتخابی اولیه (S) با مجموعه برچسب‌های L باشد. بر اساس الگوریتم (۱)، در فاز افزونگی ابتدا قدرت تاثیر هر ویژگی محاسبه می‌شود. طبق رابطه زیر، قدرت تاثیر ویژگی f_i که با λf نشان داده می‌شود عبارت است از:

$$\lambda_{fi} = Rel(f) - Red(f) \quad (22)$$

به طوری که اصطلاح اول به مقدار ارتباط ویژگی با برچسب کلاس و اصطلاح دوم به مقدار افزونگی ویژگی fi با سایر ویژگی‌های قبلاً انتخاب شده اشاره می‌کند. همان‌طور که پیش‌تر نیز گفته شد، مقدار ارتباط ویژگی عبارت است از:

$$Rel(f) = I(f; C|S) \quad (23)$$

از طرف دیگر افزونگی ویژگی fi ، شامل دو قسمت است. قسمت اول، عبارت است از افزونگی ویژگی fi با سایر ویژگی‌های قبلاً انتخاب شده، با در نظر گرفتن برچسب‌های کلاس و قسمت دوم شامل افزونگی برچسب‌ها نسبت به یکدیگر با در نظر گرفتن ویژگی fi است. با در نظر گرفتن $L^* = L$ ، می‌توان گفت افزونگی ویژگی fi از رابطه زیر به دست می‌آید:

$$Red(f) = I(f; S|L) + I(f; L|L^*) \quad (24)$$

$$I(f; S|L) = I(f; S) - I(f; S|L) \quad (25)$$

و

$$I(f; L|L^*) = I(f; L) - I(f; L|L^*) \quad (26)$$

حال با جای‌گذاری روابط (24)، (25) و (26) در رابطه شماره (22) می‌توان قدرت تاثیر ویژگی fi را از رابطه زیر به دست آورد:

$$\begin{aligned} \lambda_f &= Rel(f) - Red(f) \\ &= I(f; L) - I(f; L; S) - (I(f; S) - I(f; S; L)) - (I(f; L) - I(f; L; L^*)) \\ &= I(\{f, L\}; L^*) - I(f; S) \end{aligned} \quad (27)$$

طبق رابطه (22) و با در نظر گرفتن اندازه پارامتر $t=I$ و $n=1$ رابطه بالا به صورت زیر خلاصه می‌شود:

$$\begin{aligned} \lambda_f &= V_2(\{f, L\}_1 \times L_1^*) + V_2(\{f, S\}_1 \times L_1^*) - V_2(f \times S_1^*) \\ &= \sum_{l_i \in L^*} I(f; l_i) + \sum_{l_i \in L} \sum_{l_j \in L^*} I(l_i; l_j) \\ &\quad + \sum_{l_i \in L^*} I(f; l_i) + \sum_{f_i \in S} \sum_{l_j \in L^*} I(f_i; l_j) - \sum_{f_j \in S} I(f; f_j) \\ &= \sum_{f_j \in \{f \cup S\}} \sum_{l_i \in L^*} I(f_j; l_i) + \sum_{g \in \{f \cup L\}} \sum_{l_i \in L^*} I(g; l_i) - \sum_{f_j \in S} I(f; f_j) \end{aligned} \quad (28)$$

بعد از محاسبه میزان تاثیر هر کدام از ویژگی‌ها، الگوریتم پیشنهادی، ویژگی با قدرت تاثیر بالاتر را انتخاب و به مجموعه ویژگی‌های انتخابی از میان مجموعه ویژگی‌های انتخاب شده (S) در مجموعه جدید R اضافه می‌کند. در مرحله بعد مجدداً الگوریتم پیشنهادی میزان ارتباط مجموعه جدید R با مجموعه برچسب‌ها را چک می‌کند. میزان ارتباط ویژگی‌های مجموعه جدید R با مجموعه برچسب‌ها (L) و نیز مجموعه S با L براساس رابطه (22) به صورت زیر محاسبه می‌شود:

$$I(R;L) = V_2(R'_1 \times L'_1) = \sum_{f_j \in R} \sum_{l_j \in L} I(f_j; l_j) \quad (29)$$

$$I(S;L) = V_2(S'_1 \times L'_1) = \sum_{f_j \in S} \sum_{l_j \in L} I(f_j; l_j)$$

اگر مقدار ارتباط مجموعه انتخابی جدید (R) با مجموعه برچسب‌ها (L) از میزان ارتباط مجموعه ویژگی‌های اولیه (S) بیشتر بود، (یعنی $I(R;L) > I(S;L)$) الگوریتم متوقف می‌شود. در غیر این صورت همین روند دوباره ادامه می‌یابد. به این صورت که باز قدرت تاثیر ویژگی‌های باقی‌مانده در مجموعه S محاسبه و سپس ویژگی با بالاترین قدرت تاثیر انتخاب می‌شود. همان‌طور که گفته شد این کار تا زمانی ادامه می‌یابد که قدرت ارتباط ویژگی‌های انتخاب شده در مجموعه R و برچسب‌های L از قدرت تاثیر ویژگی‌های انتخاب شده قبلی با برچسب‌ها یکسان باشد. اما گاهی اوقات به منظور کاهش بیشتر اندازه زیر مجموعه انتخابی می‌توان شرط را به صورتی تغییر داد که زودتر متوقف شود. به عنوان مثال می‌توان شرط توقف را به صورت زیر در نظر گرفت.

$$Rel(R;L) \leq \alpha Rel(S;L) \quad (30)$$

هر چه قدر مقدار α کوچکتر باشد تعداد ویژگی‌های انتخاب شده نیز کمتر خواهد بود.

۴- نتایج آزمایشگاهی

به منظور تعیین میزان کارایی روش پیشنهادی، تعدادی آزمایش روی چندین داده آموزشی معروف و چندبرچسبی و چندین الگوریتم انتخاب ویژگی چندبرچسبی مختلف به نام $wMLDA$ [۱۶]، $MDDMf$ [۱۸]، $MDDMp$ [۱۸]، ami [۳۵] و pfs [۳۷] که اخیراً ارائه شده‌اند و جزئیات آن‌ها در بخش‌های قبلی آورده شده است، انجام شده است. خصوصیات داده‌های آموزشی چندبرچسبی شامل نام، تعداد نمونه‌های آموزشی، تعداد ویژگی‌ها و تعداد برچسب‌های هر نمونه آموزشی در جدول ۱ نشان داده شده است.

جدول ۱: مشخصات داده‌های آموزشی استفاده شده

نام داده های آموزشی	تعداد نمونه‌های هر داده	تعداد ویژگی‌های هر داده	تعداد برچسب‌های هر داده
emotions	۵۹۳	۷۲	۶
scene	۲۴۰۷	۲۹۴	۶
Nci9	۶۰	۹۷۱۲	۱

همچنین به منظور ارزیابی عملکرد روش‌های مختلف، از معیارهای مربوط به داده‌های آموزشی چندبرچسبی از جمله موارد زیر استفاده شده است. با در نظر گرفتن $Y_i \subseteq L$ به عنوان زیرمجموعه صحیح برچسب‌های نمونه آموزشی i و $Z_i \subseteq L$ به عنوان زیرمجموعه پیش‌بینی شده از برچسب‌ها آنگاه این معیارها عبارتند از:

افت همینگ^۱: افت همینگ، درصد برچسب‌هایی که به درستی دسته‌بندی نشده‌اند را محاسبه می‌کند. یعنی نمونه‌هایی که به برچسب نادرست نسبت داده شده باشد یا برچسب صحیحی که پیش‌بینی نشده باشد. مقدار این معیار از طریق رابطه زیر به دست می‌آید:

$$Hamming Loss(h, T) = \frac{1}{|T|} \sum_{i=1}^{|T|} \frac{|Y_i \Delta Z_i|}{|L|} \quad (31)$$

به طوری که Δ تفاوت متقارن بین دو مجموعه است.

یک خطا^۲: این معیار تعداد دفعاتی را می‌شمارد که برچسب با بالاترین درجه متعلق به یک نمونه، متعلق به آن نمونه نباشد.

$$OE = \frac{1}{m} \sum_{i=1}^m \mathbb{1}[\arg \max_{y \in L} f(x_i, y) \notin Y'_i] \quad (32)$$

به طوری که در این رابطه، اگر شرط درون عملگر $\mathbb{1}[\]$ برقرار باشد نتیجه یک برگردانده می‌شود. در غیر این صورت نتیجه صفر است.

پوشش^۳: 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.

¹ Hamming less

² One Error

³ Coverage

$$CV = \frac{1}{m} \sum_{i=1}^m \max_{y \in Y_i} \text{rank}(f_i, y) - 1 \quad (33)$$

افت رتبه¹: افت رتبه میانگین تعداد دفعاتی را می‌شمارد که یک برچسب مرتبط رتبه‌ای بالاتر از یک برچسب نامرتبط گرفته باشد یعنی ترتیب برچسب‌ها معکوس شده باشد.

$$RL = \frac{1}{m} \sum_{i=1}^m \frac{1}{|Y_i| |\bar{Y}_i|} \left| \{(y', y'') \mid f(f_i, y') \leq f(f_i, y''), (y', y'') \in Y_i \times \bar{Y}_i\} \right| \quad (34)$$

y'' به مجموعه متمم y' اشاره می‌کند.

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

$$AP = \frac{1}{m} \sum_{i=1}^m \frac{1}{|Y_i|} \sum_{y \in \bar{Y}_i} \frac{|\{y' \in Y_i : \text{rank}(f_i, y') \leq \text{rank}(f_i, y)\}|}{\text{rank}(f_i, y)} \quad (35)$$

هر چه قدر مقدار به‌دست‌آمده به ازای معیارهای افت همینگ، افت رتبه، یک خطا و پوشش به ازای یک الگوریتم کمتر باشد، الگوریتم کارایی بهتری دارد. از طرف دیگر هر چه قدر متوسط دقت ویژگی‌های به‌دست‌آمده به ازای یک الگوریتم بیشتر باشد، الگوریتم مطلوب‌تر است. مقادیر به‌دست‌آمده به ازای معیارهای مختلف با اجرای نسخه چندبرچسبی الگوریتم دسته بندی [40] KNN محاسبه شده است. همچنین، تمام نتایج در آزمایشات بر روی یک سیستم با 8 گیگابایت رم و کارت گرافیک RTX 3050 به‌دست آمده است.

از آنجا که در الگوریتم پیشنهادی، تعداد ویژگی‌های انتخابی از قبل مشخص نیست، پس ابتدا الگوریتم پیشنهادی را 5 بار به صورت مستقل روی هر کدام از داده‌های آموزشی اجرا می‌کنیم؛ سپس میانگین تعداد ویژگی‌های انتخاب‌شده برای هر کدام از داده‌های آموزشی را با سایر الگوریتم‌ها مقایسه می‌کنیم. پس از 5 بار اجرای الگوریتم پیشنهادی، این الگوریتم به ترتیب برای داده‌های Emotion, Scene و Nci9 به تعداد 40، 50 و 35 ویژگی انتخاب کرده است. حال هر یک از الگوریتم‌های انتخاب ویژگی دیگر را با همان تعداد ویژگی مشخص شده روی هر داده آموزشی اجرا می‌کنیم. سپس نتایج به‌دست‌آمده برای معیارهای مختلف از جمله افت همینگ، افت رتبه، پوشش، متوسط دقت و یک خطا را با هم مقایسه می‌کنیم. نتایج به‌دست‌آمده برای معیارهای مختلف و الگوریتم‌های مختلف در جدول‌های 2 تا 6 نشان داده شده است. در هر یک از این جدول‌ها، محور افقی نشان‌دهنده یک داده آموزشی و محور عمودی یک الگوریتم انتخاب ویژگی

² Ranking less

چندبرچسبی را نشان می‌دهد. همچنین مقادیر میانگین به ازای هر سه داده آموزشی و هر کدام از الگوریتم‌ها در سطر آخر آورده شده است. هر چه قدر معیارهای افت همینگ، افت رتبه، پوشش و یک خطا به ازای یک الگوریتم کمتر باشد، آن الگوریتم بهتر است. همچنین هر چه قدر معیار متوسط دقت به ازای یک الگوریتم بیشتر باشد، آن الگوریتم مطلوب‌تر است.

در جدول ۲، نتایج مربوط به معیار ارزیابی افت همینگ نمایش داده شده است. همان‌طور که قبلاً گفته شد هرچه قدر مقدار به‌دست‌آمده به ازای معیار افت همینگ در یک الگوریتم کمتر باشد، آن الگوریتم بهتر است. طبق مقادیری که جدول ۲ نشان می‌دهد، الگوریتم پیشنهادی به ازای داده آموزشی scene بهترین مقدار و به ازای داده‌های emotion و nci9 دومین بهترین مقدار را به‌دست آورده است. همچنین مقادیر متوسط آورده شده در سطر آخر، نشان می‌دهد الگوریتم پیشنهادی به‌طور متوسط روی هر سه داده آموزشی تقریباً ۱۵ درصد بهتر از سایر الگوریتم‌ها عمل کرده است.

جدول ۲: افت همینگ

	MDDMp	MDDMf	MVMD	wMLD	scls	fimf	ami	pdf	MLOSFS
emotion	0.91	0.79	0.81	0.86	0.76	0.769	0.75	0.89	0.78
scene	0.85	0.87	.89	0.79	0.84	0.75	0.78	0.69	0.67
nci9	0.99	0.89	0.78	0.97	0.68	0.78	0.84	0.89	0.69
Avg	0.916	0.85	.86	0.873	0.76	0.766	0.79	0.823	0.71

در جدول ۳ مقادیر به‌دست آمده به ازای معیار افت رتبه نشان داده شده است. هر چه قدر مقدار به‌دست آمده به ازای مقدار افت رتبه در یک الگوریتم کمتر باشد، الگوریتم کارایی بهتری دارد. طبق نتایج به‌دست‌آمده، الگوریتم پیشنهادی، الگوریتم MDDMf و الگوریتم wMLDA هر کدام به ترتیب به ازای داده‌های scene, emotions, nci9 بهترین مقادیر را به‌دست آورده اند. همان‌طور که مقدار متوسط آورده شده در سطر آخر نشان می‌دهد، می‌توان نتیجه گرفت که الگوریتم پیشنهادی بعد از الگوریتم MDDMf بر اساس معیار افت رتبه تقریباً ۶ درصد نسبت به سایر الگوریتم‌های دیگر (به جز MDDMf) برتری دارد.

جدول ۳: افت رتبه

	MDDMp	MDDMf	MVMD	wMLD	scls	fimf	ami	pdf	MLOSFS
emotion	0.37	0.19	0.29	0.34	0.225	0.31	0.186	0.34	0.3
scene	0.17	0.19	0.24	0.16	0.14	0.13	0.18	0.29	0.12
nci9	0.21	0.05	0.04	0.03	0.05	0.08	0.15	0.9	0.04
Avg	0.25	0.1433	0.19	0.176667	0.178333	0.17	0.172	0.51	0.153333

در جدول شماره ۴، نتایج به‌دست‌آمده به ازای معیار پوشش آورده شده است. همانند معیارهای افت رتبه و افت همینگ، الگوریتمی بهتر است که مقدار به‌دست‌آمده به ازای معیار پوشش، در آن کمتر باشد. طبق نتایج به‌دست‌آمده،

الگوریتم پیشنهادی اگرچه تنها به ازای داده آموزشی nci9 بهترین مقدار را به دست آورده است؛ اما متوسط مقدار به دست آمده به ازای همه داده‌های آموزشی (emotions, scene, nci9) که در سطر آخر جدول نشان داده شده است از سایر الگوریتم‌ها تقریباً ۱۹ درصد کمتر است که این عملکرد بالای الگوریتم پیشنهادی را نشان می‌دهد.

جدول ۴: معیار پوشش

	MDDMp	MDDMf	MVMD	wMLD	scls	fimf	ami	pdf	MLOSFS
emotion	10	15	17	64	58	49	84	24	15
scene	24	28	56	48	35	56	48	25	26
nci9	167	166	168	158	148	190	210	187	115
Avg	67	69.66667	80.33	90	80.333	98.33	114	78.66	57.66

در جدول شماره ۵، مقادیر مربوط به معیار متوسط دقت آورده شده است. برخلاف سایر معیارها، هر چه قدر مقدار به دست آمده به ازای این معیار در یک الگوریتم بیشتر باشد، الگوریتم کارایی بالاتری دارد. طبق مقادیر به دست آمده، بیشترین متوسط دقت به ازای داده‌های emotion, scene, nci9 به ترتیب به ازای الگوریتم fimf, scls و الگوریتم پیشنهادی بوده است. همچنین میانگین متوسط دقت‌ها روی هر سه داده آموزشی، به ازای الگوریتم scls از سایر الگوریتم‌ها بیشتر بوده است. هر چند الگوریتم پیشنهادی با اختلاف کمی از الگوریتم scls تقریباً ۷ درصد نسبت به سایر الگوریتم‌ها بهتر عمل کرده است.

جدول ۵: معیار متوسط دقت

	MDDMp	MDDMf	MVMD	wMLD	scls	fimf	ami	pdf	MLOSFS
emotion	75%	71%	69%	76%	73%	81%	82%	69%	80%
scene	64%	66%	78%	84%	89%	54%	38%	48%	54%
nci9	15%	28%	41%	32%	64%	46%	57%	67%	67%
Avg	51.3%	55%	62.6%	64%	75.3%	60.33%	59%	61.33%	64.33%

در نهایت در جدول شماره ۶، نتایج مربوط به معیار یک خطا آورده شده است. طبق نتایج به دست آمده الگوریتم پیشنهادی به ازای دو داده آموزشی emotion و nci9 و الگوریتم MDDMp به ازای داده آموزشی scene بهترین مقادیر را به دست آورده‌اند. همچنین سطر آخر نشان می‌دهد الگوریتم پیشنهادی به طور متوسط تقریباً ۱۵ درصد عملکرد بهتری نسبت به سایر الگوریتم‌ها داشته است.

جدول ۶: معیار یک خطا

	MDDMp	MDDMf	MVMD	wMLD	scls	fimf	ami	pdf	MLOSFS
emotion	1	1	0.9	0.8	1	0.79	0.98	0.79	0.67
scene	0.39	0.78	0.48	0.68	0.59	0.47	0.59	0.78	0.64
nci9	0.59	0.58	0.49	0.6	0.57	0.42	0.26	0.64	0.35
Avg	0.66	0.786	0.623	0.69	0.72	0.56	0.61	0.736	0.533

در این مقاله، یک الگوریتم انتخاب ویژگی چندبرچسبی را برای داده‌های آموزشی با جریان ویژگی که در ویژگی‌ها با گذر زمان و به تدریج به مجموعه داده‌ها اضافه می‌شوند، به نام MLOSFS، با استفاده از معیار اطلاعات متقابل چندگانه معرفی شد. این الگوریتم پیشنهادی به صورت دو فاز عمل می‌کند و با در نظر گرفتن وابستگی بین برچسب‌ها، ویژگی‌ها را در داده‌های آموزشی جریانی انتخاب می‌کند. برای ارزیابی الگوریتم پیشنهادی، از چندین الگوریتم دیگر انتخاب ویژگی چندبرچسبی و همچنین از سه داده آموزشی معروف و پرکاربرد با حجم بزرگ استفاده شده است. علاوه بر این، مقادیر به دست آمده برای پنج معیار مختلف از جمله افت همینگ، افت رتبه، متوسط دقت، یک خطا و پوشش نیز مورد ارزیابی قرار گرفته است. نتایج به دست آمده نشان می‌دهد که الگوریتم پیشنهادی روی داده‌های آموزشی استفاده شده به ازای معیارهای افت همینگ، افت رتبه، پوشش، متوسط دقت و یک خطا، به طور متوسط با اختلاف ۱۵، ۶، ۱۹، ۷ و ۱۳ درصدی عملکرد بهتری نسبت به سایر الگوریتم‌ها داشته است. هر چند به ازای معیارهای افت رتبه و متوسط دقت الگوریتم پیشنهادی با اختلاف ناچیزی که قابل چشم‌پوشی است از الگوریتم‌های MDDMf و scls در رتبه دوم قرار دارد.

مراجع

- [۱] R. E. Schapire and Y. Singer, *BoosTexter: A Boosting-based System for Text Categorization*. 2000, pp. 135-168.
- [۲] B. Wu, E. Zhong, A. Horner, and Q. Yang, "Music Emotion Recognition by Multi-label Multi-layer Multi-instance Multi-view Learning," presented at the Proceedings of the 22nd ACM international conference on Multimedia, Orlando, Florida, USA, 2014.
- [۳] W. Chen, J. Yan, B. Zhang, Z. Chen, and Q. Yang, "Document Transformation for Multi-label Feature Selection in Text Categorization," presented at the Proceedings of the 2007 Seventh IEEE International Conference on Data Mining, 2007.
- [۴] C. H. Park and M. Lee, "On applying linear discriminant analysis for multi-labeled problems," *Pattern Recognition Letters*, vol. 29, no. 7, pp. 878-887, 2008/05, doi: <https://doi.org/10.1016/j.patrec.2008.01.003>.
- [۵] M. Hatami, P. Mehrmohammadi, and P. Moradi, "A Multi-Label Feature Selection Based on Mutual Information and Ant Colony Optimization," in *2020 28th Iranian Conference on Electrical Engineering (ICEE)*, 4-6 Aug. 2020, pp. 1-6, doi: 10.1109/ICEE50131.2020.9260852.
- [۶] M. Wang, H. Li, D. Tao, K. Lu, and X. Wu, "Multimodal Graph-Based Reranking for Web Image Search %J Trans. Img. Proc.," vol. 21, no. 11, pp. 4649-4661, 2012, doi: 10.1109/tip.2012.2207397.
- [۷] W. Ding *et al.*, "Subkilometer crater discovery with boosting and transfer learning %J ACM Trans. Intell. Syst. Technol.," vol. 2, no. 4, pp. 1-22, 2011, doi: 10.1145/1989734.1989743.

- [^۸] J. Liu, Y. Lin, S. Wu, and C. Wang, "Online Multi-label Group Feature Selection," *Knowledge-Based Systems*, vol. 143, pp. 42-57, 2018.
- [^۹] J. Liu, Y. Lin, Y. Li, W. Weng, and S. Wu, "Online Multi-label Streaming Feature Selection Based on Neighborhood Rough Set," vol. 84, pp. 273-287, 2018, doi: 10.1016/j.patcog.2018.07.021.
- [^{۱۰}] J. González-López, S. Ventura, and A. Cano, "Distributed Selection of Continuous Features in Multilabel Classification Using Mutual Information," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, no. 7, pp. 228-240, 2020, doi: 10.1109/TNNLS.2019.2944298.
- [^{۱۱}] Y. Lin, Q. Hu, J. Liu, J. Li, and X. Wu, "Streaming Feature Selection for Multilabel Learning Based on Fuzzy Mutual Information," *IEEE Transactions on Fuzzy Systems*, vol. 25, no. 6, pp. 1491-1507, 2017, doi: 10.1109/TFUZZ.2017.2735947.
- [^{۱۲}] J. Liu, Y. Lin, Y. Li, W. Weng, and S. Wu, "Online multi-label streaming feature selection based on neighborhood rough set," *Pattern Recognition*, vol. 84, pp. 273-287, 2018.
- [^{۱۳}] R. A. Fisher, "THE USE OF MULTIPLE MEASUREMENTS IN TAXONOMIC PROBLEMS," *Annals Eugen*, vol. 7, pp. 179-188, 1936, doi: 10.1111/j.1469-1809.1936.tb02137.x.
- [^{۱۴}] ن. قاسم‌زاده, "استخراج ویژگی از سیگنال‌های مغزی در سیستم‌های ارتباط مغز - رایانه به منظور and س. حقی‌پور, " مجله فناوری اطلاعات در طراحی مهندسی, vol. 11, no. 2, pp. 65-78, 2019.
- [^{۱۵}] H. Wang, C. Ding, and H. Huang, "Multi-label Linear Discriminant Analysis," in *Computer Vision – ECCV 2010*, Berlin, Heidelberg, K. Daniilidis, P. Maragos, and N. Paragios, Eds., 2010// 2010: Springer Berlin Heidelberg, pp. 126-139.
- [^{۱۶}] J. Xu, "A weighted linear discriminant analysis framework for multi-label feature extraction," *Neurocomputing*, vol. 275, pp. 107-120, 2018/01/31/ 2018, doi: <https://doi.org/10.1016/j.neucom.2017.05.008>.
- [^{۱۷}] X. Lin and X.-w. Chen, "Mr.KNN: soft relevance for multi-label classification," presented at the Proceedings of the 19th ACM international conference on Information and knowledge management, Toronto, ON, Canada, 2010.
- [^{۱۸}] Y. Zhang and Z.-H. Zhou, "Multilabel dimensionality reduction via dependence maximization," *ACM Trans. Knowl. Discov. Data*, vol. 4, no. 3, pp. 1-21, 2010, doi: 10.1145/1839490.1839495.
- [^{۱۹}] A. Gretton, O. Bousquet, A. Smola, and B. Schölkopf, "Measuring Statistical Dependence with Hilbert-Schmidt Norms," in *Algorithmic Learning Theory*, Berlin, Heidelberg, S. Jain, H. U. Simon, and E. Tomita, Eds., 2005// 2005: Springer Berlin Heidelberg, pp. 63-77.
- [^{۲۰}] J. Xu, J. Liu, J. Yin, and C. Sun, "A multi-label feature extraction algorithm via maximizing feature variance and feature-label dependence simultaneously," *Knowledge-Based Systems*, vol. 98, pp. 172-184, 2016/04/15/ 2016, doi: <https://doi.org/10.1016/j.knosys.2016.01.032>.
- [^{۲۱}] L. Sun, S. Ji, and J. Ye, "Hypergraph spectral learning for multi-label classification," presented at the Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining, Las Vegas, Nevada, USA, 2008.
- [^{۲۲}] J. Xu and Z. H. Mao, "Multilabel Feature Extraction Algorithm via Maximizing Approximated and Symmetrized Normalized Cross-Covariance Operator," *IEEE Transactions on Cybernetics*, vol. 51, no. 7, pp. 3510-3523, 2021, doi: 10.1109/TCYB.2019.2909779.
- [^{۲۳}] J. Zhou, D. Foster, R. Stine, and L. Ungar, "Streaming feature selection using alpha-investing," presented at the Proceedings of the eleventh ACM SIGKDD international conference

- on Knowledge discovery in data mining, Chicago, Illinois, USA, 2005. [Online]. Available: <https://doi.org/10.1145/1081870.1081914>
- [۲۴] X. Hu, P. Zhou, P. Li, J. Wang, and X. Wu, "A survey on online feature selection with streaming features," *Frontiers of Computer Science*, vol. 12, no. 3, pp. 479-493, 2018/06/01 2018, doi: 10.1007/s11704-016-5489-3.
- [۲۵] N. Al Nuaimi, M. Masud, M. Serhani, and N. Zaki, "Streaming Feature Selection Algorithms for Big Data: A Survey," *Applied Computing and Informatics*, vol. ahead-of-print, 01/01 2019, doi: 10.1016/j.aci.2019.01.001.
- [۲۶] A. Rafie, P. Moradi, and A. Ghaderzadeh, "A Multi-Objective Online Streaming Multi-Label Feature Selection using Mutual information," *Expert Systems with Applications*, p. 119428, 2022/12/13/ 2022, doi: <https://doi.org/10.1016/j.eswa.2022.119428>.
- [۲۷] ف. احمد حمزه, "انتخاب ویژگی مبتنی بر الگوریتم جستجوی هارمونی با الگوریتم نیوی and سلیمانان قره چپق", vol. 11, no. 2, pp. 31-46, 2019. مجله فناوری اطلاعات در طراحی مهندسی.
- [۲۸] W. Ding *et al.*, "Subkilometer Crater Discovery with Boosting and Transfer Learning," *ACM TIST*, vol. 2, p. ۲۰۱۱۰۱/۰۷, ۳۹. doi: 10.1145/1989734.1989743.
- [۲۹] W. J. McGill, "Multivariate information transmission," *Psychometrika*, vol. 19, no. 2, pp. 97-116, 1954/06/01 1954, doi: 10.1007/BF02289159.
- [۳۰] L. Yu and H. Liu, "Efficient Feature Selection via Analysis of Relevance and Redundancy," *J. Mach. Learn. Res.*, vol. 5, pp. 1205–1224, 2004.
- [۳۱] S. Liang, Z. Liu, D. You, and W. Pan, "Online multi-label stream feature selection based on neighborhood rough set with missing labels," *Pattern Analysis and Applications*, vol. 25, no. 4, pp. 1025-1039, 2022/11/01 2022, doi: 10.1007/s10044-022-01067-2.
- [۳۲] J. Liu, Y. Lin, J. Du, H. Zhang, Z. Chen, and J. Zhang, "ASFS: A novel streaming feature selection for multi-label data based on neighborhood rough set," *Applied Intelligence*, vol. 53, no. 2, pp. 1707-1724, 2023.
- [۳۳] J. Lee and D.-W. Kim, "SCLS: Multi-label feature selection based on scalable criterion for large label set," *Pattern Recognition*, vol. 66, pp. 342-352, 2017/06/01/ 2017, doi: <https://doi.org/10.1016/j.patcog.2017.01.014>.
- [۳۴] J. Lee and D.-W. Kim, "Fast multi-label feature selection based on information-theoretic feature ranking," *Pattern Recognition*, vol. 48, no. 9, pp. 2761-2771, 2015/09/01/ 2015, doi: <https://doi.org/10.1016/j.patcog.2015.04.009>.
- [۳۵] J. L. H. L. D.-W. Kim, "Approximating mutual information for multi-label feature selection," *Electronics Letters*, vol. 48, no. 15, pp. 929 - 930, 30 July 2012, doi: 10.1049/el.2012.1600.
- [۳۶] F. R. K. Chung, R. L. Graham, P. Frankl, and J. B. Shearer, "Some intersection theorems for ordered sets and graphs," *Journal of Combinatorial Theory, Series A*, vol. 43, no. 1, pp. 23-37, 1986/09/01/ 1986, doi: [https://doi.org/10.1016/0097-3165\(86\)90019-1](https://doi.org/10.1016/0097-3165(86)90019-1).
- [۳۷] J. Lee and D. W. Kim, "Efficient multivariate feature filter using conditional mutual information," *Electronics Letters*, vol. 48, pp. 161-162, 02/02 2012, doi: 10.1049/el.2011.3063.
- [۳۸] W. McGill, "Multivariate information transmission," *Transactions of the IRE Professional Group on Information Theory*, vol. 4, no. 4, pp. 93-111, 1954, doi: 10.1109/TIT.1954.1057469.
- [۳۹] T. M. Cover and J. A. Thomas, *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)*. Wiley-Interscience, 2006.
- [۴۰] M.-L. Zhang and Z.-H. Zhou, "ML-KNN: A lazy learning approach to multi-label learning," *Pattern recognition*, vol. 40, no. 7, pp. 2038-2048, 2007.