دوره هجدهم، شماره پاییز ۱۴۰۴

مجله فناوری اطلاعات در طراحی مهندسی

Information Technology in Engineering Design

http://sanad.iau.ir/journal/ited

یک الگوریتم انتخاب ویژگی چند برچسبی پویا با استفاده از تحلیل مولفه اصلی

مریم رحمانی نیا*(۱)

(۱) گروه مهندسی کامپیوتر، واحد کرمانشاه، دانشگاه آزاد اسلامی، کرمانشاه، ایران*

(تاریخ دریافت: ۱۴۰۳/۱۱/۲۸ تاریخ پذیرش: ۱۴۰۴/۰۳/۱۰)

چکیده

داده‌های آموزشی چند برچسبی با مجموعه‌ای زیاد از ویژگی‌ها سروکار دارند که کارایی بسیاری از برنامه‌های کاربردی را مختل کرده است. به منظور رفع این مشکل، الگوریتم‌های انتخاب ویژگی چند برچسبی ارائه شده‌اند. با این حال، بیشتر این الگوریتم‌ها فرض می‌کنند که داده‌های آموزشی پیش از شروع الگوریتم در دسترس هستند، در حالی که داده‌ها با مرور زمان به مجموعه داده‌ها اضافه می‌شوند. اخیراً الگوریتم‌های انتخاب ویژگی چند برچسبی پویا، با انتخاب زیر مجموعه‌ای مهم در داده‌ها توجه زیادی را جلب کرده‌اند. با این وجود، هیچ کدام از این الگوریتم‌ها نتوانسته‌اند حدواسط خوبی میان معیارهای ارزیابی مختلف بدست آورند. در این مقاله قصد داریم یک روش انتخاب ویژگی چند برچسبی پویا با استفاده از تجزیه و تحلیل مولفه اصلی و نظریه اطلاعات ارائه دهیم. در این روش در ابتدا مجموعه برچسب‌ها با استفاده از اولین مولفه اصلی کاهش یافته و سپس با استفاده از نظریه اطلاعات وابستگی میان ویژگی‌ها و برچسب کاهش یافته محاسبه شده و ویژگی‌های مهم انتخاب می‌شود. برای ارزیابی روش پیشنهادی تعدادی آزمایش روی چندین داده آموزشی معروف و چندین الگوریتم انتخاب ویژگی چند برچسبی انجام شده است. نتایج نشان می‌دهد که الگوریتم پیشنهادی نتایج بهتری در مقایسه با سایر الگوریتم‌ها بدست آورده است.

کلمات کلیدی: الگوریتم های انتخاب ویژگی پویا، PCA، اطلاعات متقابل

*عهده‌دار مکاتبات:

مریم رحمانی نیا

نشانی: گروه مهندسی کامپیوتر، واحد کرمانشاه، دانشگاه آزاد اسلامی، کرمانشاه، ایران

پست الکترونیکی: ma.rahmaninia@iau.ac.ir

۱- مقدمه

امروزه اشیا چند برچسبی در دامنه‌های کاربردی مختلفی مانند طبقه‌بندی تصویر [1]، طبقه‌بندی متن [3,2]، طبقه-بندی عملکرد ژن‌ها [4] و تشخیص احساست موسیقی فراگیر شده‌اند [5]. به‌عنوان مثال یک فرد مسن ممکن است از بسیاری از بیماری‌ها از جمله دیابت، فشار خون بالا و بیماری عروق کرونر رنج ببرد یا یک فیلم ممکن است هم احساسی، هم هنری و هم ورزشی باشد [6].

همان‌گونه که داده‌های آموزشی تک برچسبی دارای تعداد زیادی از ویژگی‌ها هستند، این نوع از داده‌ها نیز دارای تعداد زیادی ویژگی هستند که بسیاری از آنها ممکن است هیچ گونه ربطی به مساله اصلی نداشته باشند و یا اینکه اضافی[1] باشند. وجود این ویژگی‌های اضافی گاهی اوقات کارایی الگوریتم‌های یادگیری ماشین را با مشکل روبرو می‌کند و باعث کاهش عملکرد آن‌ها می‌شود. گاهی اوقات هم باعث افزایش نیازهای محاسباتی و حافظه مصرفی می‌شوند. روش‌های کاهش بعد چند برچسبی، به عنوان یکی از محبوب‌ترین تکنیک‌های پیش بردازش داده‌ها با حذف ویژگی‌های اضافی و بی‌ربط کارایی این الگوریتم‌ها را افزایش می‌دهد. این روش‌ها را می‌توان در دو دسته انتخاب ویژگی چند برچسبی و استخراج ویژگی چند برچسبی دسته‌بندی کرد.

روش‌های استخراج ویژگی چند برچسبی، ابعاد ویژگی‌ها را از طریق تکنیک‌های مختلف مانند نگاشت فضا یا انتقال فضا، کاهش می‌دهند. اگرچه با اینکار ارتباطات میان ویژگی‌ها حفظ شده اما فضای ویژگی‌ها تغییر می‌کند و در کاربردهای که نیاز به معنای اصلی داده‌ها است عملا بدون استفاده هستند. نمونه‌هایی از روش‌های استخراج ویژگی چند برچسبی عبارت است از [9,7]. در اکثر این روش‌ها بعد از پیش پردازش داده‌های، مجموعه برچسب‌ها با استفاده از توابع انتقال، به یک مجموعه جدید منتقل می‌شود و سپس با استفاده از دسته بندهای چند برچسبی دسته‌بندی می‌شوند.

از طرف دیگر روش‌های کاهش بعد مبتنی بر انتخاب ویژگی با انتخاب زیر مجموعه‌ای از ویژگی‌های مهم و کارآمد از میان مجموعه ویژگی‌های اولیه، بدون هیچگونه تغییری در داده‌های آموزشی، ابعاد آنها را کاهش می‌دهند. از این رو، در روش‌های انتخاب ویژگی چند برچسبی به خوبی معنای فیزیکی ویژگی‌های اصلی باقی مانده و با توجه به خوانایی

[1] Redundant

آن‌ها دارای بسیاری از مزیت‌ها هستند. این روش‌ها را می‌توان بطور کلی به روش‌های فیلتر[1]، پوششی[2] و تعبیه شده[3] دسته‌بندی کرد.

روش‌های اتخاب ویژگی فیلتر بدون استفاده از هیچ گونه دسته‌بندی به انتخاب ویژگی‌ها می‌پردازند. در این روش‌ها در ابتدای کار ویژگی‌ها براساس خصوصیت ذاتی خودشان رتبه‌دهی شده سپس، k تا از بهترین آن‌ها انتخاب شده و به این صورت ابعاد داده‌های آموزشی کاهش پیدا می‌کند [11,10]. در روش‌های انتخاب ویژگی پوششی، با استفاده از یک دسته‌بند از پیش تعیین شده به ارزیابی زیر مجموعه‌های انتخابی از ویژگی‌های می‌پردازد. به منظور انتخاب زیر مجموعه‌های مختلف استراتژی‌های مختلفی وجود دارد. می‌توان از روش‌های جلو رونده[4]، عقبگرد[5] و یا دو سویه استفاده کرد [12].

روش‌های تعبیه شده با به حداقل رساندن خطا با استفاده از الگوریتم‌های یادگیری ماشین، سعی دارند ویژگی‌های کارامدتری را انتخاب کنند. این روش‌ها هم از مزایای روش‌های فیلتر و هم از مزایای روش‌های پوششی استفاده می‌کنند [13]. همچنین به منظور حذف ویژگی‌های نامربوط و اضافی، از ماتریس‌های انتخاب ویژگی و معیارهای تنظیم‌سازی استفاده می‌کنند. برخی از روش‌هایی که در این دسته قرار می‌گیرند عبارتند از [15,14].

اگرچه روش‌های ارائه شده در بالا دارای مزایای زیادی هستند اما هنگامی که با داده‌های جریانی و پویا روبرو می‌شوند عملکرد خود را از دست می‌دهند. در همه این الگوریتم‌ها لازم است قبل از شروع الگوریتم همه داده‌های آموزشی در دسترس باشد. هر چند در دنیای واقعی بدست آوردن همه داده‌های آموزشی قبل از شروع الگوریتم غیر عملی است. به عنوان مثال، عناوین داغ و مهم روز بروز در شبکه‌های اجتماعی در حال تغییر و اضافه شدن است. هنگامی که یک عنوان داغ و مهم در شبکه‌های اجتماعی ظاهر می شود، کلمات کلیدی زیادی نیز به همراه آن به مخزن شبکه‌های اجتماعی نیز اضافه می‌شود. به عبارتی این کلمات همان ویژگی‌هایی به حساب می‌آیند که عنوان ها با آن‌ها شناخته می‌شود.

در الگوریتم‌های انتخاب ویژگی پویا چند برچسبی فرض بر این است که داده‌ها بصورت پشت سر هم و یکی‌یکی به مجموعه داده‌ها اضافه می‌شوند. همچنین مجموعه ویژگی‌های انتخابی عبارت است از یک زیر مجموعه از مجموعه

[1] Filter
[2] Wrapper
[3] Embedded
[4] Forward
[5] backward

ویژگی‌های تا آن لحظه دیده شده که نسبت به سایر ویژگی‌ها اهمیت بیشتری دارند [18,16]. بطور کلی سه شرط اصلی در رابطه با این الگوریتم‌ها باید در نظر گرفته شود:

۱. اول اینکه این الگوریتم به هیچ دانش پیش زمینه‌ایی نیاز ندارند.
۲. دوم اینکه عملیات بروز رسانی مجموعه ویژگی‌های انتخاب شده باید بصورت کارآمد و سریع اتفاق بیافتد.
۳. سوم اینکه الگوریتم‌های طبقه‌بندی باید از دقت کافی به واسطه انتخاب این ویژگی‌ها برخوردار باشند.

در این مقاله، به منظور حل چالش‌های موجود در انتخاب ویژگی‌های چند برچسبی، از تحلیل مولفه‌های اصلی (PCA) استفاده شده است. تحلیل مولفه‌های اصلی به دلیل توضیح واریانس بالا و سادگی در کاهش ابعاد به صورت خطی به عنوان یک روش برتر شناخته می‌شود. این روش می‌تواند نویز را کاهش دهد، تجزیه و تحلیل چند متغیره را تسهیل کند و به تجزیه و تحلیل بصری داده‌ها کمک کند. همچنین، محاسبات PCA معمولاً سریع و کارآمد است. به همین دلایل، از این روش کاهش ابعاد در مقاله حاضر بهره گرفته شده است.

در این مقاله یک الگوریتم انتخاب ویژگی جندبرچسبی پویا در داده‌های آموزشی جریانی ارائه می‌دهیم که با استفاده از رویکرد تجزیه و تحلیل مولفه اصلی[۱] که یک روش استخراج ویژگی است به انتخاب ویژگی‌های جریانی می‌پردازد. در ابتدا به منظور نگاشت مجموعه برچسب‌ها به یک برچسب از تجزیه و تحلیل مولفه اصلی استفاده کرده و با استفاده از اولین مولفه اصلی بدست آمده برای مجموعه برچسب‌های کاهش یافته به انتخاب ویژگی‌های مهم و کارآمد می‌پردازیم. در ادامه ابتدا پیشینه تحقیق و پیش زمینه که شامل مروری بر نظریه اطلاعات و PCA است معرفی می‌شود. سپس الگوریتم پیشنهادی و نتایج آزمایشگاهی در بخش‌های سوم و چهارم آورده شده است. درنهایت نتیجه‌گیری و کارهای آینده را در بخش پنجم معرفی می‌کنیم.

۲- کارهای مرتبط

بطور کلی روش‌های کاهش بعد داده‌های آموزشی را می‌توان در دو دسته استخراج ویژگی چند برچسبی و انتخاب ویژگی چند برچسبی قرار داد. روش‌های استخراج ویژگی چندبرچسبی با استفاده از روش‌های نگاشت فضای ویژگی یا روش‌های انتقال فضا، ابعاد داده‌های آموزشی را کاهش می‌دهند. یکی از مشهورترین روش‌های کاهش بعد برمبنای استخراج ویژگی، روش تجزیه و تحلیل تبعیض آمیز خطی[۲] (LDA) است که هدف آن، پیدا کردن یک تابع نگاشت

[۱] Principle component analysis
[۲] linear discriminative analysis

خطی است بطوریکه شباهت درون کلاسی را بشینه و در آن واحد شباهت میان کلاس‌ها را کمینه کند [19]. اخیرا LDA، به چند روش به صورت چند برچسبی[۱] (MLDA) گسترش داده شده است [21,20]. تفاوت اصلی این روش‌ها به نحوه وزن‌دهی برچسب‌ها برمی‌گردد. بر این اساس روش wMLDAb [22]، از وزن‌دهی دودویی، wMLDAe از وزن‌دهی براساس آنتروپی [23]، wMLDAc از وزن‌دهی براساس همبستگی[۲] [20]، wMLDAf از وزن‌دهی بر اساس منطق فازی[۳] [24] و wMLDAd از وزن‌دهی بر اساس وابستگی[۴] [21] استفاده کرده است. در [25]، یک روش تحت ناظر به منظور کاهش بعد بر اساس بیشینه‌سازی وابستگی ارائه شده است که به‌نام MDDM معروف است. MDDM تلاش می‌کند تا با استفاده از معیار استقلال هیلبرت اشمیت [26] وابستگی میان ویژگی‌ها و برچسب‌ها را بیشینه کند. این راهکار به دو صورت به‌نام MDDMp و MDDMf ارائه شده است. روش MDDMp، براساس جهت‌های نگاشت متعامد[۵] عمل می‌کند در حالی که روش MDDMf ویژگی‌های نگاشت شده را به صورت متعامد کاهش می‌دهد. در [27]، نویسندگان نشان دادند که روش MDDMp را می‌توان به صورت کمترین مربعات خطا نشان داد. آنها همچنین، با اضافه کردن PCA به تابع هدف روش MDDMp یک روش جدید بنام MVMD[۶] ارائه دادند. در [28]، یک روش انتخاب ویژگی با استفاده از ابرگراف[۷] براساس داده‌ها و برچسب‌ها برای نشان دادن وابستگی برچسب‌های مختلف ارائه شده است. در این روش از فرمول یادگیری طیفی ابرگراف[۸] به منظور دسته بندی چندبرچسبی استفاده شده است. در [23]، یک الگوریتم انتساب برچسب براساس مفهوم آنتروپی ارائه شده است که در آن به ازای برچسب‌های مختلف بر اساس مفهوم آنتروپی یک وزن به نمونه‌های آموزشی چندبرچسبی داده شده است. این روش که به ELA[۹] معروف است هر نمونه را براساس تعداد برچسب‌هایی که دارد کپی می‌کند. سپس معکوس تعداد برچسب-های آن را به عنوان وزن به هر کدام از نمونه‌ها اختصاص می‌دهد. طبق این الگوریتم نمونه‌هایی که دارای برچسب بیشتری هستد وزن کمتری نیز می‌گیرند و از چرخه الگوریتم‌های یادگیری حذف می‌شوند. روش مجموعه چند برچسبی یا LP[۱۰] در [29] به بازیابی اطلاعات موزیک یا بطور خاص‌تر به استخراج احساسات مختلف از موزیک‌ها اعمال شده

[۱] Multi label linear discriminative analysis
[۲] Correlation
[۳] Fuzy
[۴] Dependency
[۵] orthonormal projection directions
[۶] Multi-label feature extraction via maximizing feature variance and feature-label dependence simultaneously
[۷] hypergraph
[۸] hypergraph spectral learning
[۹] Entropy-based Label Assignment
[۱۰] Label Powerset

است. در این روش، داده‌های چند برچسبی با نگاشت مجموعه برچسب‌های هر نمونه به یک کلاس به نمونه‌های تک برچسبی تبدیل می‌شوند. تعداد کل کلاس‌ها برابر با تعداد کل مجموعه‌های مجزا از برچسب‌ها است. اما اگر چه روش LP به خوبی ارتباط بین برچسب‌ها را به منظور انتخاب ویژگی بررسی می‌کند اما این روش از مشکل تعداد کلاس‌ها رنج می‌برد [30]. در [31]، الگوریتم انتقال مساله برشی بنام PPT[1] به منظور بهبود کارایی روش LP ارائه شد. این روش، الگوهای با تعداد برچسب‌های اندک را با در نظر گرفتن مقدار آستانه τ به راحتی از مجموعه حذف می‌کند. در [32] یک روش انتخاب ویژگی با تابع انتقال با استفاده از PPT به منظور بهبود دسته‌بندی داده‌های تصاویر و ژن‌ها ارائه شد. در این روش در ابتدا، داده‌های آموزشی چند برچسبی با استفاده از روش PPT انتقال یافته سپس یک الگوریتم جلو رونده ترتیبی با استفاده از مفهوم اطلاعات متقابل به آن اعمال شده است. نتایج بدست آمده نشان می‌دهد، این روش در مقایسه با $\mathrm{PPT} + \mathrm{X}^2$ نتایج بهتری تولید می‌کند. یکی از محدودیت‌های اصلی روش‌های ذکر شده استفاده از تابع انتقال به منظور محاسبه ارتباط و وابستگی در میان ویژگی‌های انتخاب شده و برچسب‌ها است.

در مقابل، روش‌های انتخاب ویژگی چند برچسبی، تنها با انتخاب زیرمجموعه‌ای از میان ویژگی‌های موجود و با حفظ ساختار داده‌ها، باعث کاهش ابعاد داده‌های آموزشی می‌شوند [36,33].

اگر چه این روش‌ها به صورت کارآمد به فرایند انتخاب ویژگی در داده‌های چند برچسبی می‌پردازند و قادر به رفع مشکل ازدحام ابعاد هستند، قادر به رفع مشکل جریانی بودن داده‌های آموزشی نبودند و هنگامی که داده‌ها بصورت تدریجی کامل می‌شدند نمی‌توانستند ابعاد آنها را کاهش دهند.

روش‌های انتخاب ویژگی برخط چند برچسبی، فرض می‌کنند که ویژگی‌ها با گذر زمان به داده‌های آموزشی با بیش از یک برچسب اضافه می‌شود. [37] یک روش انتخاب ویژگی برخط با جریان ویژگی چندبرچسبی است که با استفاده از نظریه همسایگی مجموعه‌های سخت[2] به انتخاب ویژگی‌های مرتبط و غیر افزونه می‌پردازد. [16] نیز با استفاده نظریه همسایگی اطلاعات متقابل[3] در طی دو فاز به انتخاب ویژگی‌های مهم در داده‌های آموزشی برخط با جریان ویژگی چند برچسبی می‌پردازد. در فاز اول، از میان گروه‌های تا آن لحظه اضافه شده، گروه‌های موثر را شناسایی و سپس در فاز دوم، به انتخاب ویژگی‌های مهم و حذف ویژگی‌های افزونه از هر گروه می‌پردازد.

در [16] یک الگوریتم انخاب ویژگی چند برچسبی پویا به منظور انتخاب ویژگی‌های مهم در داده‌های چند برچسبی ارائه شده است. روش آنها شامل دو فاز فاز درون گروهی و فاز بین گروهی است. در فاز اول، اهمیت ویژگی‌های تازه

[1] pruned problem transformation
[2] neighborhood rough set theory
[3] Neighborhood Mutual Information

وارد محاسبه شده و در فاز بین گروهی افزونگی ویژگی‌ها ارزیابی شده و ویژگی‌های اضافی از مجموعه حذف می‌شوند. دو روش دیگر در [17] به منظور فرآیند انتخاب ویژگی در داده‌های آموزشی چند برچسبی پویا با استفاده از اطلاعات متقابل فازی ارائه شده است. نویسندگان در [16] یک الگوریتم انتخاب ویژگی چند برچسبی پویا براساس مجموعه همسایگی سخت به منظور شناسایی یک زیرمجموعه مرتبط از ویژگی‌ها ارائه داده اند. این روش، مجموعه همسایگی تک برچسبی در داده‌های آموزشی تک برچسبی را به منظور انجام فرآیند انتخاب ویژگی در داده‌های آموزشی چند برچسبی گسترش داده است. هچنین مشکل انتخاب دانه‌بندی در مجموعه همسایگی سخت را با استفاده از بیشینه کردن نزدیک‌ترین همسایه هر نمونه برای دانه‌بندی حل کرده است. مجموعه داده‌های سخت تنها برای داده‌های آموزشی گسسته کاربرد دارد. همچنین پیچیدگی روش‌هایی که براساس این نظریه عمل می‌کنند بسیار بالا است. یک روش انتخاب ویژگی دیگر در [38] به منظور انتخاب ویژگی در داده‌های آموزشی چند برچسبی ارائه شده است. این روش مبتنی بر دانه‌بندی طیفی و اطلاعات متقابل است و برچسب‌ها را به مجموعه‌ایی با ابعاد کاهش یافته برای گرفتن همسایگی‌ها تبدیل می‌کند. همچنین هنگامی که یک گروه از ویژگی‌ها اضافه می‌شود فرآیند انتخاب یا حذف انجام می‌شود. در این روش نیز، هنگامی که گروه تازه وارد از ویژگی‌ها مرتبط یا اضافی است خوب عمل نمی‌کند. یک روش انتخاب ویژگی چند برچسبی پویا دیگر در [39] ارائه شده است. این روش از یک رویکرد فیلتر به منظور انتخاب ویژگی‌های ضروری طی سه فاز استفاده می‌کند. در فاز اول از یک چهار چوب چندهدفه بر اساس بهینه‌سازی ازدحام ذرات روی یک گروه تازه وارد از ویژگی‌ها عمل می‌کند. در مرحله دوم نیز، مقدار افزونگی ویژگی‌های تازه وارد با مجموعه ویژگی‌های قبلا انتخاب شده محاسبه می‌شود. در فاز سوم، ویژگی‌هایی را از لیست ویژگی‌های انتخاب شده قبلی که دیگر نسبت به ویژگی‌های تازه معرفی شده مرتبط نیستند را پیدا کرده و آنها را کنار می‌گذارد. در این روش از یک رویکرد چند هدفه به منظور بهبود کارایی الگوریتم استفاده شده است. بطور کلی می‌توان گفت همه روش‌های کاهش بعد ارائه شده دارای یک سری مزایا هستند اما هیچ کدام از این روش‌ها نتوانسته‌اند به خوبی همه چالش‌های مربوط به داده‌های برخط را مرتفع کنند.

از یک طرف اگر چه روش‌های استخراج ویژگی چند برچسبی با انتقال مجوعه فضای ویژگی‌ها به یک فضای جدید وابستگی‌ها را حفظ می‌کنند اما معنای اولیه ویژگی ها از بین می‌رود. همچنین این روش‌ها به یک سری اطلاعات از پیش تعیین شده در مورد مساله نیاز دارند. از طرف دیگر، روش‌های انتخاب ویژگی معنای اولیه ویژگی‌ها را حفظ می‌کنند ولی برای بدست آوردن وابستگی میان ویژگی‌ها به یک سری محاسبات طولانی نیازمند هستند.

در این مقاله قصد داریم از مزایای هر دو روش استفاده کنیم و یک روش انتخاب ویژگی برخط در جریان داده‌های چند برچسبی ارائه دهیم. در روش ارائه شده در ابتدا مجموعه فضای برچسب‌ها را به یک برچسب انتقال می‌دهیم و

سپس به فرایند انتخاب ویژگی‌های مهم در مجموعه داده‌های جریانی می‌پردازیم. برای اینکار از نظریه اطلاعات به منظور شناسایی ویژگی های مهم استفاده می‌کنیم.

۲-۱- نظریه اطلاعات متقابل

در نظریه احتمالات و نظریه اطلاعات، اطلاعات متقابل بین دو متغیر تصادفی معیاری برای نشان دادن میزان وابستگی متقابل آن دو متغیر می‌باشد. به بیان دیگر در حقیقت این معیار «میزان اطلاعات» به دست آمده (مثلاً در واحد بیت) در مورد یک متغیر تصادفی از طریق متغیر تصادفی دیگر را نشان می‌دهد. مفهوم اطلاعات متقابل ذاتاً مرتبط با آنتروپی یک متغیر تصادفی که میزان اطلاعات موجود در یک متغیر تصادفی را نشان می‌دهد، می‌باشد. اطلاعات متقابل میزان شباهت بین توزیع مشترک p(X,Y) و ضرب احتمال‌های حاشیه ای یعنی p(X)p(Y) را مشخص می‌سازد.

تعریف ۱: اطلاعات متقابل بین دو متغیر تصادفی X و Y را به صورت زیر می‌توان تعریف نمود:

$$I(X;Y) = \sum_{y \in Y} \sum_{x \in X} p(x,y) \log \frac{p(x,y)}{p(x)\,p(y)} \tag{1}$$

که در رابطه فوق p (x , y) تابع توزیع احتمال مشترک X و Y و p (x) و p (y) تابع‌های توزیع احتمال حاشیه ای می‌باشند.

در صورتی که متغیرهای تصادفی پیوسته باشند، رابطه به صورت زیر بر اساس انتگرال معین دوگانه تعریف می‌گردد:

$$I(X;Y) = \int_{Y} \int_{X} p(x,y) \log\left(\frac{p(x,y)}{p(x)\,p(y)} \right) dx\, dy \tag{2}$$

در رابطه فوق اکنون p (x , y) تابع چگالی احتمال مشترک X و Y ، و p (x) و p (y) تابع‌های چگالی احتمال حاشیه‌ای به ترتیب X و Y می‌باشند. اگر لگاریتم در پایه ۲ استفاده شود، واحد اطلاعات متقابل بیت خواهد بود. اطلاعات متقابل میزان اطلاعاتی که بین X و Y مشترک است را اندازه می‌گیرد.

۲-۲- تجزیه و تحلیل مولفه اصلی

تحلیل مولفه اساسی، روشی برای استخراج متغیرهای مهم (به شکل مولفه) از مجموعه بزرگی متغیرهای موجود در یک مجموعه داده است. تحلیل مولفه اساسی در واقع یک مجموعه با بُعد پایین از ویژگی‌ها را از یک مجموعه دارای بُعد بالا استخراج می‌کند تا به ثبت اطلاعات بیشتر با تعداد کمتری از متغیرها کمک کند.

یک مولفه اساسی یک ترکیب خطی نرمال شده از پیش‌بین‌های اصلی موجود در مجموعه داده است. فرض می‌شود یک مجموعه از پیش‌بین‌ها به صورت $X^1, X^2...,X^p$ وجود دارد.

مولفه‌های اساسی این مجموعه از پیش‌بین‌ها را می‌توان بدین شکل نوشت:

$$Z^1 = \Phi^{11}X^1 + \Phi^{21}X^2 + ... + \Phi^{p1}X^p \quad (3)$$

که در آن Z^1 اولین مولفه اساسی و Φ^{p1} بردار بار شامل بردارهای $(\Phi^1, \Phi^2...)$ اولین مولفه اساسی هستند. بنابراین اولین مولفه اساسی، یک ترکیب خطی از پیش‌بین‌های اصلی است که بیشترین واریانس موجود در مجموعه داده‌ها را در بر می‌گیرد.

۳- الگوریتم پیشنهادی

دراین بخش یک روش انتخاب ویژگی چند برچسبی با استفاده از اطلاعات متقابل و PCA ارائه می‌دهیم. در روش پیشنهادی در ابتدا، مجموعه برچسب‌ها به یک برچسب با استفاده از PCA کاهش پیدا می‌کند. سپس مقدار ارتباط هر کدام از ویژگی‌ها با در نظر گرفتن برچسب کاهش یافته محاسبه می‌شود. از آنجا که نظریه اطلاعات متقابل می‌تواند وابستگی‌های خطی و غیرخطی را شناسایی کند پس از این معیار به منظور محاسبه میزان ارتباط یک ویژگی با برچسب کلاس استفاده می شود. حال اگر ویژگی با برچسب کلاس مرتبط بود الگوریتم وارد فاز جستجوی ویژگی‌های اضافی می‌شود و با حذف ویژگی‌های اضافی اندازه مجموعه ویژگی‌های انتخابی را کاهش می‌دهد. در ادامه دو معیار ارتباط و افزونگی جهت تعیین میزان ارتباط و افزونگی ویژگی‌ها معرفی می‌شود که در الگوریتم پیشنهادی از آن‌ها استفاده می‌شود.

تعریف ۱ (ارتباط ویژگی): میزان ارتباط و تاثیر یک ویژگی $F=[f_1, f_2,..., f_n]$ با مجموعه برچسب‌های $L=[l_1, l_2,..., l_n]$ با استفاده از رابطه زیر بدست می‌آید:

$$\mathrm{Re}l(F;C) = \sum_{f_i \in F} \sum_{c \in C} p(f_i, c) \log \frac{p(f_i, c)}{p(f_i)p(c)} \quad (4)$$

بطوریکه I(F,C) به اطلاعات متقابل میان ویژگی F و برچسب کاهش یافته C=PCA(L) اشاره می‌کند.

تعریف ۲ (افزونگی ویژگی): ویژگی F_2 نسبت به ویژگی F_1 اضافی در نظر گرفته می‌شود اگر و تنها اگر

$$I(F_1;F_2) > I(F_1;C) \ , \ I(F_2;C) < I(F_1;C) \quad (5)$$

در ادامه فازهای الگوریتم پیشنهادی را با جزییات بیشتر آورده‌ایم.

فاز اول: این فاز به ازای همه نمونه‌های آموزشی و تنها یک بار اجرا می‌شود. از آنجا که هر نمونه آموزشی دارای چند برچسب است، پس در این فاز مجموعه برچسب‌های هر نمونه آموزشی $L=[l_1,l_2,...,l_n]$ با استفاده از PCA به اولین مولفه اصلی نگاشت می‌شود. بطوریکه وابستگی و ارتباط میان برچسب‌ها حفظ شده و تنها معنای فیزیک آن‌ها تغییر پیدا کرده است. برچسب کاهش پیدا کرده C نامیده می‌شود.

فاز دوم: این فاز به ازای ورود هر ویژگی جدید F اجرا می شود و ارتباط ویژگی تازه وارد با برچسب جدید C با استفاده از تعریف ۱ محاسبه می‌شود (خط ۷ تا ۱۰ شکل ۱). اگر مقدار اطلاعات متقابل میان برچسب نگاشت شده جدید C و ویژگی F بزرگتر از صفر بود آنگاه می‌توان نتیجه گرفت که ویژگی مربوطه با مجموعه برچسب‌ها مرتبط است و در مجموعه ویژگی‌های انتخاب شده بنام SF قرار می‌گیرد. در غیر اینصورت ویژگی F با مجموعه برچسب‌ها مرتبط نیست. اگرچه انتخاب مقدار صفر به منظور مقدار آستانه در کاربردهای واقعی کمی غیر ممکن است. یک راه‌حل ایمن استفاده از یک مقدار مثبت و غیر صفر α است.

```
1  Input: Feature stream F, Label space L and relevancy threshold alpha;
2  Output: selected features SF
3  Converted the label set L into one label C based on PCA;
4  SF = { } ;
5  Repeat
6     Get feature F
7   /*checking relevance based on definition 2*/
8   If Rel (F, C) < alpha
9      Insert F in SF;
10   End
11 /*checking redundancy based on definition 3*/
12   For each feature Fi in SF
13 If Rel (Fi, C)>Rel (F, C) && I (F, Fi)>I (F, C)
14   SF=SF\ Fi;
15   Break;
16  End
17 Until no new features or stopping criteria met;
18 Return SF
```

شکل ۱: الگوریتم پیشنهادی

فاز سوم: هنگامی که یک ویژگی با برچسب کلاس مرتبط بود، این فاز اجرا می‌شود و ویژگی‌های افزونه شناسایی می‌شوند. به منظور تعیین میزان افزونگی یک ویژگی از تعریف ۲ استفاده می‌شود. این فاز به ازای تمام ویژگی‌های قبلا انتخاب شده F_i موجود در مجموعه SF اجرا می‌شود. در صورتی که وابستگی ویژگی $F_i \in SF$ با برچسب C بیشتر از وابستگی ویژگی تازه وارد F و برچسب C باشد و نیز وابستگی میان دو ویژگی F_i و F بیشتر از میزان ارتباط ویژگی F_i باشد، آنگاه ویژگی F_i از مجموعه SF حذف می‌شود (خط ۱۳تا ۱۶ شکل ۱). به منظور وضوح بیشتر شبه کد الگوریتم پیشنهادی در شکل ۱ نشان داده شده است.

۴- نتایج آزمایشگاهی

به منظور تعیین میزان کارایی روش پیشنهادی، تعدادی آزمایش روی چندین داده آموزشی معروف و چند برچسبی به‌نام yeast, flags, emotions، scene[۱]، bibtex و birds و چندین الگوریتم انتخاب ویژگی چند برچسبی برخط مختلف به‌نام GRRO [40]، fimf [35]، pmu [41] و d2f [42] که اخیرا ارائه شده‌اند انجام شده است. خصوصیات داده‌های آموزشی چند برچسبی در جدول ۱ نشان داه شده است. همچنین به منظور ارزیابی از معیارهای زمان اجرا، تعداد ویژگی‌های انتخاب شده، افت همینگ[۲]، پوشش[۳] و متوسط دقت[۴] استفاده شده است [10]. هر چقدر مقدار بدست آمده به ازای معیار های زمان اجرا، تعداد ویژگی های انتخاب شده، افت همینگ و پوشش به ازای یک الگوریتم کمتر باشد الگوریتم کارایی بهتری دارد. از طرف دیگر هر چقدر دقت ویژگی‌های بدست آمده به ازای یک الگوریتم بیشتر باشد الگوریتم مطلوب‌تر است. مقادیر بدست آمده به ازای معیارهای مختلف با ا جرای نسخه چند برچسبی الگوریتم دسته بند ML_KNN [43] است.

جدول ۱: خصوصیات داده‌های آموزشی

دامنه	#برچسب	#ویژگی	#نمونه ها
Biology	14	103	2417
image	7	19	194
Yeast			

[۱] https://www.uco.es/kdis/mllresources/

[۲] Hamming Loss

[۳] Coverage

[۴] Average precision

emotions	593	72	6	Music
scene	2407	294	6	Image
bibtex	7395	1836	159	text
birds	645	260	19	audio

در ادامه نتایج بدست آمده به ازای معیارهای مختلف آورده شده است. در هر جدول، ستون ها نشان دهنده یک داده آموزشی، سطره نشان دهنده یک الگوریتم انتخاب ویژگی چند برچسبی است. همچنین بهترین مقادیر در هر ستون به صورت درشت و زیر خط‌دار نشان داده شده است.

۴-۱- تعداد ویژگی‌های انتخاب شده

در جدول ۲ نتایج بدست آمده مربوط به تعداد ویژگی‌های انتخابی به ازای داده‌های آموزشی مختلف و الگوریتم‌های انتخاب ویژگی مختلف نشان داده شده است. طبق نتایج بدست آمده الگوریتم پیشنهادی به ازای همه داده‌های آموزشی به جز داده آموزشی bibtex از همه الگوریتم‌ها تعداد کمتری ویژگی انتخاب کرده است. به عنوان مثال همان‌طور که جدول ۲ نشان می‌دهد الگوریتم پیشنهادی بطور متوسط ۱ ویژگی به ازای داده‌های آموزشی yeast و flags انتخاب کرده در حالی که GRRO، fimf، pmu و d2f بترتیب ۱۰۳، ۱۰۳، ۵ و ۴ ویژگی را انتخاب کرده‌اند. این نتایج نشان می‌دهد الگوریتم پیشنهادی معمولا تعداد کمتری ویژگی نسبت به سایر ویژگی‌ها انتخاب کرده است و کارایی بالای آن را نشان می‌دهد.

جدول ۲: نتایج بدست آمده به ازای تعداد ویژگی‌های انتخاب شده

	yeast	flags	emotons	scene	bibtex	birds
GRRO	103	19	72	89	18	24
fimf	103	19	78	8	10	11
pmu	5	5	5	5	27	15
d2f	4	4	4	4	16	9
proposed method	1	1	3	4	15	7

۴-۲- افت همینگ

نتایج مربوط به معیار افت همینگ در جدول ۳ نشان داده شده است. همانطور که جدول نشان می‌دهد الگوریتم پیشنهادی به ازای داده‌های آموزشی flags و bibtex و birds کمترین مقدار را بدست آورده است. این در حالی است

که الگوریتم های GRRO و d2f در هیچ کدام و الگوریتم pmu در داده yeast و الگوریتم fimf در دو داده آموزشی emotions و scene بهترین مقادیر را بدست آورده‌اند. بطور کلی می‌توان گفت الگوریتم پیشنهادی بطور متوسط عملکرد بهتری از لحاظ معیار افت همینگ داشته است.

جدول ۳: نتایج بدست آمده به ازای معیار افت همینگ

	yeast	flags	emotons	scene	bibtex	birds
GRRO	0.9859	0.8484	0.9513	0.9999	0.9710	0.9186
fimf	0.8894	0.6725	0.8531	0.9671	0.9922	0.9995
pmu	0.8743	0.6593	0.8853	0.9975	0.9745	0.8942
d2f	0.8765	0.7055	0.8804	0.9698	0.9913	0.7845
proposed method	0.8812	0.5925	0.9059	0.9965	0.9477	0.7515

۴-۳- پوشش

در جدول ۴ نتایج بدست آمده به ازای معیار پوشش نشان داده شده است. طبق نتایج بدست آمده الگوریتم پیشنهادی به ازای داده های yeast و flags، الگوریتم pmu به ازای داده آموزشی bibtex، الگوریتم d2f به ازای داده‌های آموزشی emotion، scene و birds بهترین مقادیر را بدست آورده‌اند. از این میان الگوریتم‌های fimf و GRRO در هیچ موردی بهترین مقادر را بدست نیاورده‌اند.

جدول ۴ : نتایج بدست آمده به ازای معیار پوشش

	yeast	flags	emotons	scene	bibtex	birds
GRRO	905.14	63.28	200	931.8	13.19	5.84
fimf	7.36	3.66	2.37	2.04	9.70	9.75
pmu	7.88	3.86	2.66	2.05	4.28	4.54
d2f	7.61	3.81	2.53	2.03	8.77	1.65
proposed method	6.84	3.75	2.72	2.33	4.80	2.11

۴-۴- متوسط دقت

در جدول ۵ نتایج مربوط به معیار متوسط دقت نشان داده شده است. طبق نتایج بدست آمده، الگوریتم پیشنهادی بطور متوسط به ازای سه داده آموزشی، الگوریتم fimf به ازای ۲ داده آموزشی و الگوریتم d2f به ازای تنها یک داده آموزشی بهترین متوسط دقت را بدست آورده‌اند.

جدول ۵ : نتایج بدست آمده به ازای معیار متوسط دقت

	yeast	flags	emotons	scene	bibtex	birds
GRRO	30.97	46.14	34.51	24.66	81.54	91.52

fimf	68.33	80.52	69.97	47.72	79.54	78.66
pmu	64.61	82.5	67.35	50.66	46.57	68.58
d2f	66.25	80.29	68.72	50.82	78.62	92.53
proposed method	63.45	89.14	66.52	45.78	87.50	94.50

۴-۵- زمان اجرا

متوسط زمان اجرا به ازای اجراهای مختلف روی داده‌های آموزشی مختلف در جدول ۶ نشان داده شده است. همان‌طور که جدول نشان می‌دهد الگوریتم fimf به ازای داده‌های آموزشی flags، emotions، bibtex و Birds همواره در زمان کمتری ویژگی‌ها را انتخاب کرده است. بعد از این الگوریتم نیز الگوریتم پیشنهادی قرار دارد که به ازای داده های آموزشی yeast و scene زمان کمتری سپری کرده است.

جدول ۶: نتایج بدست آمده به ازای معیار زمان اجرا

	yeast	flags	emotons	scene	bibtex	birds
GRRO	1.33	19.51	0.53	6.88	45.5	48.1
fimf	0.082	0.05	0.05	0.06	0.12	0.09
pmu	8.4	0.3	1.31	8.17	0.78	0.84
d2f	1.75	0.14	0.47	5.33	0.64	0.41
proposed method	0.245	0.44	0.29	0.98	0.81	0.45

۴-۶- تجزیه و تحلیل الگوریتم پیشنهادی

به منظور نشان دادن بیشتر کارایی الگوریتم پیشنهادی در این بخش متوسط مقادیر بدست آمده به ازای معیارهای تعداد ویژگی‌های انتخاب شده، زمان سپری شده، افت همینگ، پوشش و متوسط دقت آورده شده است. همانطور که در جدول شماره ۷ نشان داده شده است الگوریتم پیشنهادی بطور کلی به ازای معیارهای تعداد ویژگی‌های انتخاب شده، افت همینگ و پوشش مقادیر بهتری نسبت به سایر الگوریتم‌ها به دست آورده است. بعد از الگوریتم پیشنهادی، الگوریتم fimf به ازای زمان سپری شده بطور متوسط روی همه داده های آموزشی بهترین مقدار را بدست آورده است.

جدول ۷: متوسط مقادیر معیارهای تعداد ویژگی های انتخاب شده، زمان، افت همینگ، پوشش و متوسط دقت

	#feat	time	hamig loss	covergae	percision
GRRO	126.16	20.3	0.944	353.1	51.52
fimf	88.5	0.075	0.893	5.805	70.77
pmu	10.33	3.3	0.879	4.191	63.35

d2f	6.833	1.458	0.867	4.38	72.86
Proposed method	5.666	0.535	0.845	3.756	74.47

۵- نتیجه‌گیری

در این مقاله یک الگوریتم انتخاب ویژگی چند برچسبی درجریان داده های آموزشی چند برچسبی که به تدریج و با گذر زمان به مجموعه اضافه می‌شوند ارائه داده شد. در الگوریتم پیشنهادی در ابتدا، مجموعه برچسب ها به ازای هر نمونه آموزشی به یک برچسب با استفاده از تابع PCA انجام شد. سپس به منظور تعیین میزان افزونگی و ارتباط ویژگی از نظریه اطلاعات متقابل استفاده شد. نتایج بدست آمده به ازای داده‌های آموزشی و الگوریتم‌های انتخاب ویژگی مختلف نشان داد که الگوریتم پیشنهادی در مقایسه با سایر الگوریتم‌ها کارایی بالاتری دارد. درکارهای بعدی قصد داریم الگوریتم پیشنهادی را به گونه ایی تغییر دهیم که بتواند در جریان داده‌های گروهی نیز به فرایند انتخاب ویژگی‌های مهم بپردازد. در جریان داده‌های گروهی داده‌ها بصورت گروهی و مجتمع به مجموعه اضافه می‌شوند.

منابع

[1] م. امین طوسی, "تمام متصل به تمام پیچشی: پلی به گذشته," مجله علمی رایانش نرم و فناوری اطلاعات، جلد ۱۱، شماره ۱، ۱۴۰۲.

[2] ا. هاشمی, م. دولتشاهی and ,ح. نظام آبادی پور, "انتخاب ویژگی شورایی مبتنی بر حداقل افزونگی، حداکثر همبستگی: یک رویکرد دو هدفه بر اساس مفهوم غلبه پارتو," مجله علمی رایانش نرم و فناوری اطلاعات، جلد ۱۲، شماره ۱، سال ۱۴۰۲.

[3] ا. قیطاسی,ع. رضائی, س. ش. فلاحیه حمیدپور and ,ف. خواجه خلیلی, "ارائه یک سیستم تشخیص کامپیوتری هوشمند جدید برای تشخیص سرطان سینه با استفاده از تصاویر حرارتی," مجله علمی رایانش نرم و فناوری اطلاعات، جلد ۱۱، شماره ۴، ۱۴۰۲.

[4] A. Elisseeff and J. Weston, "A kernel method for multi-labelled classification," *Advances in neural information processing systems*, vol. 14, 2001.

[5] K. Trohidis, G. Tsoumakas, G. Kalliris, and I. P. Vlahavas, "Multi-label classification of music into emotions," in *ISMIR*, 2008, vol. 8, pp. 325-330.

[6] W. Siblini, P. Kuntz, and F. Meyer, "A review on dimensionality reduction for multi-label classification," *IEEE Transactions on Knowledge and Data Engineering*, vol. 33, no. 3, pp. 839-857, 2019.

[7] Y. Zhang and Z.-H. Zhou, "Multilabel dimensionality reduction via dependence maximization," *ACM Transactions on Knowledge Discovery from Data (TKDD)*, vol. 4, no. 3, pp. 1-21, 2010.

[8] H. Hotelling, "Relations between two sets of variates," *Breakthroughs in statistics: methodology and distribution*, pp. 162-190, 1992.

[9] K. Yu, S. Yu, and V. Tresp, "Multi-label informed latent semantic indexing," in *Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval*, 2005, pp. 258-265.

[10] R. Shaikh, M. Rafi, N. A. Mahoto, A. Sulaiman, and A. Shaikh, "A filter-based feature selection approach in multilabel classification," *Machine Learning: Science and Technology*, vol. 4, no. 4, p. 045018, 2023/10/27 2023, doi: 10.1088/2632-2153/ad035d.

[11] M. Rahmaninia and P. Moradi, "OSFSMI: Online stream feature selection method based on mutual information," *Applied Soft Computing*, vol. 68, pp. 733-746, 2018/07/01/ 2018, doi: https://doi.org/10.1016/j.asoc.2017.08.034.

[12] U. M. Khaire and R. Dhanalakshmi, "Stability of feature selection algorithm: A review," *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 4, pp. 1060-1073, 2022/04/01/ 2022, doi: https://doi.org/10.1016/j.jksuci.2019.06.012.

[13] L.-b. Qiao, B.-f. Zhang, J.-s. Su, and X.-c. Lu, "A systematic review of structured sparse learning," *Frontiers of Information Technology & Electronic Engineering*, vol. 18, pp. 445-463, 2017.

[14] L. Yuan, J. Liu, and J. Ye, "Efficient methods for overlapping group lasso," *Advances in neural information processing systems*, vol. 24, 2011.

[15] R. Tibshirani, M. Saunders, S. Rosset, J. Zhu, and K. Knight, "Sparsity and smoothness via the fused lasso," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 67, no. 1, pp. 91-108, 2005.

[16] J. Liu, Y. Lin, S. Wu, and C. Wang, "Online multi-label group feature selection," *Knowledge-Based Systems*, vol. 143, pp. 42-57, 2018.

[17] 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.

[18] P. Zhou, N. Wang, and S. Zhao, "Online group streaming feature selection considering feature interaction," *Knowledge-Based Systems*, vol. 226, p. 107157, 2021/08/17/ 2021, doi: https://doi.org/10.1016/j.knosys.2021.107157.

[19] 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.

[20] 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.

[21] 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.

[22] 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, 2008/05/01/ 2008, doi: https://doi.org/10.1016/j.patrec.2008.01.003.

[23] 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.

[24] 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.

[25] 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.

[26] 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.

[27] 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.

[28] 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.

[29] K. Trochidis, G. Tsoumakas, G. Kalliris, and I. Vlahavas, *Multi-label classification of music into emotions*. 2008, pp. 325-330.

[30] G. Tsoumakas, I. Katakis, and I. Vlahavas, *Random k-Labelsets for Multi-Label Classification*. 2011, pp. 1079-1089.

[31] J. Read, *A pruned problem transformation method for multi-label classification*. 2008.

[32] G. Doquire and M. Verleysen, "Feature selection for multi-label classification problems," presented at the Proceedings of the 11th international conference on Artificial neural networks conference on Advances in computational intelligence - Volume Part I, Torremolinos-Mÿlaga, Spain, 2011.

[33] 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.

[34] 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.

[35] 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.

[36] 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.

[37] 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.

[38] H. Wang, D. Yu, Y. Li, Z. Li, and G. Wang, "Multi-label online streaming feature selection based on spectral granulation and mutual information," in *Rough Sets: International Joint Conference, IJCRS 2018, Quy Nhon, Vietnam, August 20-24, 2018, Proceedings 6*, 2018: Springer, pp. 215-228.

[39] D. Paul, A. Jain, S. Saha, and J. Mathew, "Multi-objective PSO based online feature selection for multi-label classification," *Knowledge-Based Systems*, vol. 222, p. 106966, 2021.

[40] J. Zhang *et al.*, "Fast multilabel feature selection via global relevance and redundancy optimization," *IEEE Transactions on Neural Networks and Learning Systems*, 2022.

[41] J. Lee and D.-W. Kim, "Feature selection for multi-label classification using multivariate mutual information," *Pattern Recognition Letters*, vol. 34, no. 3, pp. 349-357, 2013.

[42] J. Lee and D.-W. Kim, "Mutual Information-based multi-label feature selection using interaction information," *Expert Systems with Applications*, vol. 42, no. 4, pp. 2013-2025, 2015/03/01/ 2015, doi: https://doi.org/10.1016/j.eswa.2014.09.063.

[43] A. Huang, R. Xu, Y. Chen, and M. Guo, "Research on multi-label user classification of social media based on ML-KNN algorithm," *Technological Forecasting and Social Change*, vol. 188, p. 122271, 2023/03/01/ 2023, doi: https://doi.org/10.1016/j.techfore.2022.122271.