Journal of Applied And Basic Machine Intelligence Research (ABMIR)

پژوهش‌های نظری و کاربردی هوش ماشینی

سال اول، شماره دوم، پاییز و زمستان ۱۴۰۲، صفحات۲۸ - ۱۶

معرفی یک روش انتخاب ویژگی چند-برچسبه مبتنی بر کمینه کردن افزونگی

حکیمه حسنیه [۱]، بهزاد ابراهیمی [۲]، فرید صابری موحد [۳]، مهدی افتخاری [۴*]

[۱] کارشناس ارشدگروه ریاضی‌کاربردی، دانشکده علوم و فناوری‌های نوین، دانشگاه تحصیلات تکمیلی صنعتی و فناوری پیشرفته، کرمان، ایران

[۲] کارشناس ارشد بخش مهندسی کامپیوتر، دانشگاه شهید باهنر کرمان، ایران

[۳] استادیارگروه ریاضی کاربردی، دانشکده علوم و فناوری‌های نوین، دانشگاه تحصیلات تکمیلی صنعتی و فناوری پیشرفته، کرمان، ایران

[۴] دانشیار بخش مهندسی کامپیوتر، دانشگاه شهید باهنر کرمان، ایران

چکیده

روش‌های انتخاب ویژگی ابزاری کارا در بهبود فرآیند یادگیری شناخته می‌شوند. هدف از یک روش انتخاب ویژگی، شناسایی ویژگی‌های مرتبط و حذف ویژگی‌های غیرمرتبط به‌منظور به دست آوردن یک زیرمجموعه مناسب از ویژگی‌ها است، به‌طوری‌که افزونگی بین ویژگی‌های انتخاب‌شده کمینه گردد. در داده‌های چند-برچسبه، این امکان وجود دارد که در صورت وجود همبستگی بین ویژگی‌ها، مقدار افزونگی در مجموعه ویژگی‌ها افزایش یابد. وجود افزونگی بین ویژگی‌ها به همراه چالش ابعاد بالای داده‌های چند-برچسبه، می‌تواند باعث افزایش حجم محاسبات، کاهش دقت و درنهایت افزایش احتمال رخ دادن خطا در پیش‌بینی و طبقه‌بندی داده‌های چند-برچسبه شود. در این مقاله، باهدف کمینه کردن افزونگی ویژگی‌های انتخابی، یک الگوریتم انتخاب ویژگی چند-برچسبه با در نظرگرفتن مدل رگرسیون کمترین مربعات خطا و تنظیم تنکی پیشنهادشده است. درانتها، با استفاده از تعدادی مجموعه داده چند-برچسبه مشهور، کارایی روش پیشنهادی بررسی می‌گردد و نتایج به‌دست‌آمده با چند روش انتخاب ویژگی چند-برچسبه متداول مقایسه می‌شود.

مقاله پژوهشی

تاریخ دریافت:
۱۴۰۱/۶/۲۹

تاریخ پذیرش:
۱۴۰۲/۳/۲۳

کلیدواژه‌ها:
انتخاب ویژگی چند-برچسبه، کمینه کردن افزونگی، رگرسیون، کمترین مربعات خطا، تنظیم تنکی

نویسنده‌مسئول:
m.eftekhari@uk.ac.ir

doi : 10.22034/ABMIR.2023.18967.1016

E-ISSN: 2821-2037 /© 2023. Published by Yazd University This is an open access article under the CC BY 4.0 License (https://creativecommons.org/licenses/by/4.0/).

۱- مقدمه

در روش‌های یادگیری با نظارت سنتی، هر نمونه تنها با یک برچسب نشان داده می‌شد تا وابستگی مفهومی کلاس را نشان دهد. بااین‌حال، در بسیاری از مسائل کاربردی در دنیای واقعی، یک نمونه ممکن است توسط مجموعه‌ای از برچسب‌ها نمایش داده شود. به‌عنوان‌مثال، در طبقه‌بندی متون [۱]، یک سند ممکن است به موضوعات مختلفی نظیر دولت و سلامت متعلق باشد. در طبقه‌بندی و حاشیه‌نویسی تصاویر [۲]، یک تصویر ممکن است دارای چندین طبقه معنایی مانند ساحل و مناطق شهری باشد. در ژنومیک عملکردی متون[۳]، هر ژن ممکن است مجموعه‌ای از کلاس‌های اساسی مانند متابولیسم، رونویسی و سنتز پروتئین را به خود اختصاص دهد. بنابراین، در مسائل انتخاب ویژگی، دو نمونه از برچسب‌گذاری موردتوجه قرار می‌گیرد: روش تک-برچسبه و روش چند-برچسبه[٤]. درروش تک-برچسبه، هر نمونه فقط با یک برچسب علامت‌گذاری می‌شود، درحالی‌که درروش چند-برچسبه، یک نمونه معمولاً با بیش از یک برچسب علامت‌گذاری می‌شود، به این معنی که هر نمونه، مجموعه‌ای از برچسب‌های متفاوت را دارا است[٥]. مشابه با داده‌های تک-برچسبه، داده‌های چند-برچسبه نیز با چالش حجیم بودن داده‌ها سروکار دارند. به‌ویژه آنکه، داده‌های چند-برچسبه با ابعاد بالا، اغلب دارای ویژگی‌های غیرضروری و زائد می‌باشند که منجر به افزایش خطا در پیش‌بینی می‌شود. برای جلوگیری از این مشکل و کاهش تاثیرات منفی ناشی از آن بر فرآیند طبقه‌بندی، باید از بین مجموعه ویژگی‌های اولیه، یک زیرمجموعه از ویژگی‌ها را به‌گونه‌ای انتخاب کرد که بیشترین تأثیر را بر روی مسئله‌دارند[٦]. در این راستا، روش‌های انتخاب ویژگی ابزاری کارا در بهبود فرآیند یادگیری شناخته می‌شوند. به‌طورکلی، هدف از یک روش انتخاب ویژگی، شناسایی یک زیرمجموعه مناسب از ویژگی‌ها و حذف ویژگی‌های غیرمرتبط است، بطوریکه افزونگی بین ویژگی‌های انتخاب‌شده کمینه گردد، و همچنین این زیرمجموعه منتخب قادر باشد تا مجموعه داده اصلی را به‌خوبی تشریح نماید [۷]. برای حل مسائل یادگیری چند-برچسبه با استفاده از روش انتخاب ویژگی، نیاز به تکنیک‌هایی است که قادر به پیدا کردن ویژگی‌های متمایز بین چند برچسب مشترک باشند. در سال‌های اخیر، دو مفهوم مهم و کلیدی تحت عناوین «تنظیم تنکی[۱]» و کمینه کردن «نرخ افزونگی[۲]» نقشی اساسی را در روش‌های انتخاب ویژگی تک-برچسبه و چند-برچسبه ایفا کرده‌اند[۸،۹]. برای مثال، لین[۳] و همکاران[۱۰]یک روش انتخاب ویژگی چند-برچسبه مبتنی بر دو مفهوم کمینه افزونگی و بیشینه ارتباط معرفی کردند. در این روش، مقیاسی جهت محاسبه اهمیت هریک از ویژگی‌ها با استفاده از کشف قابلیت آن‌ها در بهبود عملکرد یادگیری چند-برچسبه معرفی شد. تنظیم تنکی به این منظور استفاده می‌گردد تا در مسئله طراحی‌شده برای انتخاب ویژگی، بتوان نقش ویژگی‌های مؤثر را برجسته‌تر کرد. از طرف دیگر، کمینه کردن نرخ افزونگی ازاین‌جهت حائز اهمیت است که کمک می‌کند تا افزونگی بین ویژگی‌های انتخاب‌شده به میزان کافی کاهش یابد[۷]. علاوه بر این موارد، برخی از مفاهیم مانند اطلاعات مشترک[٤] و آنتروپی شانون[٥] در سال‌های اخیر توجه بسیاری از پژوهشگران را در یادگیری چند-برچسبه به خود جلب کرده‌اند. در مسائل انتخاب ویژگی، آنتروپی شانون یک روش اندازه‌گیری متغیرهای تصادفی با عدم قطعیت است و به‌طور گسترده‌ای در بخش انتخاب ویژگی چند-برچسبه، یادگیری قانون[٦] و همچنین کشف دانش‌های مختلف مورداستفاده قرارگرفته است[۱۱]. در ارتباط با مفهوم اطلاعات مشترک، لی و کیم[۷] [۱۲]، یک روش با استفاده از اطلاعات مشترک چندمتغیره برای طبقه‌بندی انتخاب ویژگی چند-برچسبه پیشنهاد و ارائه کردند. در این روش، زیرمجموعه‌ای شامل ویژگی‌های مؤثر به‌وسیله بیشینه کردن وابستگی بین ویژگی‌های منتخب و برچسب‌ها، انتخاب می‌گردد. لی و کیم[۸] [۱۳]، همچنین یک الگوریتم با استفاده از اطلاعات تعاملی برای انتخاب ویژگی چند-برچسبه مبتنی بر اطلاعات مشترک معرفی کردند. در ادامه، همان نویسندگان یک روش بهتر برای انتخاب ویژگی چند-برچسبه ارائه

[1] sparsity regularization
[2] redundancy rate
[3] Lin
[4] mutual information
[5] Shannons entropy
[6] rule learning
[7] Lee and Kim
[8] Lee and Kim

کردند که در آن با بکاربردن یک‌رویه دارای مزیت، زیرمجموعه‌های ویژگی را در خلال جستجوی ژنتیک دوباره تعریف می‌نماید [۱۴]. داکوئر و ورلیسن[1] یک الگوریتم انتخاب ویژگی بر پایه اطلاعات مشترک برای مسائل انتخاب ویژگی چند-برچسبه با استفاده از PPT برای تغییر شکل مسئله و سپس استراتژی جستجوی حریصانه[2] مبتنی بر اطلاعات مشترک هدایت‌شده به‌منظور انتخاب ویژگی‌های دارای ارتباط هرچه بیشتر، طراحی کردند [۱۵]. در سالیان گذشته، روش‌های انتخاب ویژگی چند-برچسبه موثری معرفی‌شده‌اند. برای مثال، جیان[3] و همکاران [۱۶] یک قالب انتخاب ویژگی آگاهانه چند-برچسبه جدید پیشنهاد کردند که در آن برچسبی را استخراج کند تا از میان چندین برچسب، ویژگی‌های متمایزکننده را برگزینند. آن‌ها اطلاعات مربوط به فضای چند-برچسبه را به یک فضای کم بُعد تجزیه کرده و سپس این فضای کم بُعد به خدمت گرفته‌شده را در اختیار یک الگوریتم انتخاب ویژگی گذاشتند. کای و ژو[4] [۹] یک روش جدید برای انتخاب ویژگی چند-برچسبه با استفاده از مدل «رگرسیون کمترین مربعات»[5] و تلفیقی از «یادگیری منیفلد ویژگی»[6] و «تنظیم تنکی» را موردمطالعه قراردادند. در این راستا، از مفهوم گراف تنظیم برای پیدا کردن ساختار هندسی ویژگی‌ها به‌منظور رسیدن به یک ماتریس ضریب رگرسیون بهتر استفاده‌شده است تا اهمیت هرکدام از ویژگی‌های مختلف منعکس شود. هوانگ[7] و همکاران [۱۷] یک روش انتخاب ویژگی جدید تحت عنوان امتیاز لاپلاسین مقید مبتنی بر منیفلد ارائه کردند. در این روش، از یادگیری منیفلد برای تبدیل فضای برچسب اصلی به فضای برچسب اقلیدسی و تشابه بین نمونه‌هایی که محدود به تناظر عددی برچسب‌ها شده‌اند، استفاده‌شده است. هو[8] و همکاران [۱۸] یک روش بر پایه منظم‌سازی گراف دوگانه، یعنی تنظیم گراف ویژگی و تنظیم گراف برچسب طراحی کردند. منظم‌سازی گراف ویژگی، در حفظ همبستگی ساختار هندسی ویژگی‌ها و منظم‌سازی گراف برچسب، برای کشف همبستگی بین برچسب‌ها مورداستفاده قرارگرفته است.

بااین‌حال، تحقیقات جزئی درباره بررسی تأثیر افزونگی بین ویژگی‌ها گزارش‌شده است. در یادگیری چند-برچسبه، افزونگی نه‌تنها شامل اطلاعات مشترک بین ویژگی موردنظر و ویژگی‌های انتخاب‌شده است، بلکه باید اطلاعات مشترک بین ویژگی موردنظر و همه برچسب‌ها را درزمانی که ویژگی‌های منتخب داده می‌شوند نیز دربر بگیرد.

هدف از این مقاله، بررسی تأثیر کمینه کردن افزونگی بین ویژگی‌ها بر روند انتخاب ویژگی چند-برچسبه است. در این مقاله، در گام اول، با استفاده از مسئله کمینه‌سازی کمترین مربعات خطا، مسئله انتخاب ویژگی به‌صورت یک مسئله تجزیه ماتریسی همراه با جمله تنظیم تنکی و با استفاده از نرم فروبنیوس بیان می‌شود. در گام دوم، با پیدا کردن رابطه بین ویژگی‌ها بر اساس نرخ افزونگی، سعی در کمینه کردن نرخ افزونگی و باهدف یافتن ویژگی‌های بااهمیت بیشتر می‌شود. سپس، یک تابع هدف برای مسئله بهینه‌سازی موردنظر طراحی‌شده، و یک الگوریتم بروز رسانی تکراری نیز برای حل تابع هدف متناظر با مسئله انتخاب ویژگی مطرح‌شده، موردبررسی قرار خواهد گرفت. در انتها، نتایج به‌دست‌آمده از این روش با چند روش انتخاب ویژگی چند-برچسبه مشهور مقایسه شده و کارایی روش پیشنهادی نسبت به روش‌های متداول پیشین بررسی می‌گردد.

ساختار این مقاله به شرح زیر است. در بخش دوم، نمادهای موردنیاز این مقاله آورده شده است. در بخش ۳، به تعریف مفهوم

$$F_1 = \{[1\ 0\ 0], [0\ 1\ 0], [1\ 1\ 0]\}$$
$$F^2 = \{[1\ 0\ 0], [0\ 1\ 0], [0\ 0\ 1]\}$$

در مجموعه F_1بردار $[1\ 1\ 0]$می‌تواند یک بردار زائد تلقی شود، به این مفهوم که این بردار نسبت به دو بردار دیگر این مجموعه وابسته خطی است. از طرف دیگر، مجموعهF_2، یک مجموعه نرخ افزونگی با ذکر مثال، تعریف و قضیه پرداخته‌شده است. در بخش ٤، مدل‌سازی مسئله انتخاب ویژگی چند-برچسبه و در بخش ۵، الگوریتم بهینه‌سازی متناظر و حل آن آورده شده است. همچنین در بخش ٦ نتایج عددی و تجزیه تحلیل نتایج حاصل‌شده، ذکرشده

[1] Doquire and Verleysen
[2] greedy search
[3] Jian
[4] Cai and Zhu
[5] least square regression model
[6] feature manifold learning
[7] Huang
[8] Hu

است. درنهایت، در بخش ۷ نتیجه حاصل از این مقاله را خواهیم داشت.

۲- نمادهای موردنیاز

در ابتدا نمادهای مورداستفاده در مقاله ذکر خواهد شد. در سراسر این مقاله، فرض می‌شود که مجموعه داده‌های چند-برچسبه، شامل n نمونه $x_1, x_2, \dots, x_n$ است. به‌عبارت‌دیگر، ماتریس نمونه‌ها عبارت است از:

$$X = [x_1; x_2; \dots; x_n] \in \mathbb{R}^{n\times d}$$

هم‌چنین، ماتریس برچسب‌ها به‌صورت زیر نمایش داده می‌شود:

$$Y = [y_1; y_2; \dots; y_n] \in \{0,1\}^{n\times l}$$

که هر

$$y_i = [y_i^1, y_i^2, \dots, y_i^l] \in \{0,1\}^l$$
$$i = 1, \dots, n.$$

بردار برچسب‌های متناظر با نمونه x_iاست. همچنین، 1_k ماتریسی k×k است که تمام درایه‌های آن ۱ است.

۳- نرخ افزونگی

نرخ افزونگی یک مجموعه از ویژگی‌ها یکی از مفاهیم اساسی در بحث انتخاب ویژگی محسوب می‌شود که در حقیقت بیانگر میزان تشابه بین ویژگی‌ها است. برای توضیح بیشتر ابتدا با یک مثال شروع می‌کنیم.

مثال ۳-۱. دو مجموعه زیر را در نظر بگیرید:

مستقل خطی است و اعضای آن برخلاف مجموعه F_1 از «تشابه کمتری» نسبت به یکدیگر برخوردار هستند. بنابراین، مجموعه F_2 در مقایسه با مجموعه F_1 از افزونگی کمتری برخوردار است.

یکی از راه‌های تعریف نرخ افزونگی، استفاده از مفهوم آماری «ضریب همبستگی پیرسون» است که در ادامه به آن می‌پردازیم.

تعریف ۳-۲. فرض کنید

$$F_I = \{F_{I_1}, F_{I_2}, \dots, F_{I_k}\}$$

یک زیرمجموعه از ویژگی‌ها باشد. نرخ افزونگی F_I به‌صورت زیر تعریف می‌شود:

$$RED(I) = \frac{1}{k^2}\sum_{i.j\in I}\rho(f_i, f_j) \quad (۱)$$

که $I = \{I_1, \dots, I_k\}$ و $\rho(f_i, f_j)$ ضریب همبستگی پیرسون بین دو ویژگی f_i و f_j است و عبارت است از:

$$\rho(f_i, f_j) = \frac{cov(f_i, f_j)}{\sqrt{D(f_i)D(f_j)}} \quad (۲)$$

که در آن $cov(f_i, f_j)$ کوواریانس ویژگی‌های f_i و f_j است و به‌صورت زیر تعریف می‌شود:

$$cov(f_i, f_j) = \langle f_i, f_j\rangle = f_i^T f_j \quad (۳)$$

به‌علاوه، $D(f_i)$ و $D(f_j)$ به ترتیب انحراف معیار[1] مربوط به f_i و f_j می‌باشند.

اگر داده‌ها نرمال‌سازی شوند و فرض شود که همه ویژگی‌ها دارای توزیع نرمال هستند، آنگاه ضریب همبستگی پیرسون به‌صورت زیر ساده می‌شود:

$$\rho(f_i, f_j) = cov(f_i, f_j) \quad (۴)$$

مثال ۳-۳. فرض کنید که مجموعه داده عبارت است از:

$$X = \{[1\ 0\ 0], [0\ 1\ 0], [1\ 1\ 0], [0\ 0\ 1]\}$$

در این مثال، هدف محاسبه نرخ افزونگی دو مجموعه ویژگی F_1 و F_2 است که در مثال ۳-۱ معرفی شدند. برای این منظور، داریم:

$$RED(I_1) = 0.89\ , RED(I_2) = 0.33$$

که $I_1 = \{1,2,3\}$ و $I_2 = \{1,2,4\}$. بنابراین، میزان افزونگی مجموعه ویژگی F_2 از میزان افزونگی مجموعه F_1 کمتر است.در ادامه، یک فرم ماتریسی برای نرخ افزونگی برای یک مجموعه از ویژگی‌ها بیان می‌شود.

قضیه ۳-۴. فرض کنید که ماتریس داده عبارت است از

$$X = [f_1, f_2, \dots, f_d] \in \mathbb{R}^{n\times d}$$

همچنین، فرض کنید F_I یک مجموعه از ویژگی‌ها باشد که $|I| = k$ در این صورت یک ماتریس $W \in \mathbb{R}^{d\times k}$ وجود دارد که

$$RED(\mathrm{I}) = \frac{1}{k^2}Tr(W^T X^T X W 1) \quad (۵)$$

که 1 ماتریسی $k \times k$ است که تمام درایه‌های آن ۱ است.

اثبات: فرض کنید $F_I = \{F_{I_1}, F_{I_2}, \dots, F_{I_k}\}$ یک زیرمجموعه از ویژگی‌ها باشد. از آنجاکه

$$F_I \subseteq span\{f_1, f_2, \dots, f_d\}$$

بنابراین، یک ماتریس $W \in \mathbb{R}^{d\times k}$ وجود دارد به طوری‌که

[1] Standard Deviation

$$XW = \{f_{I_1}, f_{I_2}, \dots, f_{I_k}\}$$

درنتیجه، خواهیم داشت:

$$\frac{1}{k^2} Tr(W^T X^T XW1) = \sum_{i.j \in I} \langle f_i, f_j \rangle$$

که اثبات قضیه را کامل می‌سازد.

٤- مدل‌سازی مسئله انتخاب ویژگی چند-برچسبه

در این بخش، به‌منظور ارائه یک روش انتخاب ویژگی چند-برچسبه، مدلی پیشنهاد می‌شود که از سه مفهوم زیر در ساختار آن استفاده‌شده است

۱) مسئله رگرسیون کمترین مربعات.

هدف از یک مسئله رگرسیون کمترین مربعات، کمینه‌سازی تابع هدف زیر است:

$$\min_{W \in \mathbb{R}^{d \times l}} \|XW - Y\|_F^2 \quad (٦)$$

که در آن $W \in \mathbb{R}^{d \times l}$ ماتریس ضریب (یا ماتریس وزن) رگرسیون است.

۲) کمینه کردن نرخ افزونگی

منظور از مفهوم «کمینه کردن افزونگی» در فرآیند انتخاب ویژگی، دستیابی به یک زیرمجموعه از ویژگی‌ها است که کمترین وابستگی ممکن بین ویژگی‌های آن وجود داشته باشد. به‌عبارت‌دیگر، هرچقدر مقدار نرخ افزونگی کمتر باشد، آنگاه انتظار داریم مجموعه ویژگی‌های منتخب دارای وابستگی کمتری نسبت به یکدیگر باشند. بنابراین، با توجه فرم ماتریسی نرخ افزونگی، که در رابطه (٥) بیان شد، هدف از کمینه کردن نرخ افزونگی در بحث انتخاب ویژگی عبارت است از:

$$\min_{W \in \mathbb{R}^{d \times l}} Tr(W^T X^T XW1) \quad (٧)$$

۳) تنک‌سازی ماتریس وزن W.

مقدار $\|w_i\|_2$ نشان‌دهنده اهمیت ویژگی i ام است و هرچه مقدار $\|w_i\|$ بزرگ‌تر باشد، ویژگی i ام از اهمیت بیشتری نسبت به سایر ویژگی‌ها برخوردار است. بنابراین، در بحث انتخاب ویژگی، انتظار می‌رود که فقط تعداد کمی از $\|w_i\|_2$ های محاسبه‌شده غیر صفر باشند. درنتیجه، این نیاز احساس می‌شود که ماتریس وزن W دارای خاصیت تنک بودن باشد. به همین دلیل، بحث تنک‌سازی ماتریس W در مدل انتخاب ویژگی از اهمیت زیادی برخوردار است و به‌صورت کمینه کردن یک نرم مناسب از W انجام می‌گیرد. در بسیاری از روش‌های موجود، از نرم $l_{2.1}$ به‌منظور تنک‌سازی ماتریس وزن W استفاده شد. به‌علاوه اینکه اگر $D \in \mathbb{R}^{d \times d}$ یک ماتریس قطری باشد که i امین درایه روی قطر اصلی آن به‌صورت زیر باشد:

$$D_{ii} = \frac{1}{2\|w_i\|_2}$$

در این صورت:

$$\|W\|_{2.1} = Tr(W^T DW) \quad (٨)$$

نکته قابل‌توجه این است که ممکن است حالتی رخ دهد که در آن برخی از درایه‌های قطری ماتریس D صفر گردند و یا اینکه حتی اعدادی کوچک باشند. در این حالت، در عمل به‌منظور جلوگیری از خطاهای عددی، مقایسه با یک مقدار اختلال کوچک به میزان ϵ در درایه‌های قطری ماتریس D به‌صورت زیر اعمال می‌شود:

$$D_{ii} = \frac{1}{2 \times max(\|w_i\|_2 . \epsilon)}$$

بنابراین، همان‌طور که به‌وضوح دیده می‌شود، پارامتر ϵ می‌تواند نقشی اساسی در ساختار $\|W\|_{2.1}$ ایفا کند. از طرف دیگر، بیایید نرم فروبنیوس ماتریس W را در نظر بگیریم، یعنی $\|W\|_F$. ازیک‌طرف داریم:

$$\|W\|_F \leq \|W\|_{2.1}$$

و از طرف دیگر، برتری بزرگی که $\|W\|_F$ به $\|W\|_{2.1}$ دارد این است که ساختار $\|W\|_F$ به هیچ پارامتری وابسته نیست. همین موضوع، انگیزه‌ای برای ما است و تنک‌سازی ماتریس وزن W در بحث انتخاب ویژگی را به‌صورت کمینه کردن تابع هدف $\|W\|_F$ در نظر بگیریم، یعنی،

$$\min_{W \in \mathbb{R}^{d \times l}} \|W\|_F \quad (٩)$$

اکنون، با توجه به عبارات (٧)، (٨) و (٩)، مسئله انتخاب ویژگی چند-برچسبه بر اساس مسئله رگرسیون کمترین مربعات و با استفاده از مفاهیم کمینه کردن نرخ افزونگی و تنک‌سازی ماتریس وزن W با استفاده از نرم فروبنیوس، به‌صورت مساله بهینه‌سازی زیر بیان می‌شود:

$$\min_{W\in\mathbb{R}^{d\times l}} \frac{1}{2}\|W^TX-Y\|_F^2+\frac{\alpha}{2}\|W\|_F^2+\frac{\beta}{2}Tr(W^TX^TXW1) \quad (۱۰)$$

که در آن، αپارامتر منظم‌ساز متناظر با تنک‌سازی ماتریس وزن W و β پارامتر منظم‌ساز متناظر با نرخ افزونگی است.

۵- الگوریتم بهینه‌سازی و حل آن

به‌منظور حل مسئله کمینه‌سازی (۵) تابع $f(W)$ به‌صورت زیر تعریف می‌شود:

$$f(W)=\frac{1}{2}\|XW-Y\|_F^2+\frac{\alpha}{2}\|W\|_F^2+\frac{\beta}{2}Tr(W^TX^TXW1_k). \quad (۱۱)$$

که مقدار $\frac{1}{2}$ قبل از عبارات تابع هدف به‌منظور ساده شدن با ضریب ۲ ایجادشده پس از مشتق‌گیری در نظر گرفته‌شده است. با مشتق گرفتن از عبارت (۱۱) نسبت به W و مساوی صفر قرار دادن آن خواهیم داشت:

$$\frac{\partial f}{\partial W}=X^TXW-X^TY+\alpha W+\beta X^TXW1_k=0$$

با استفاده از رابطه فوق، مقدار W به‌صورت زیر به دست خواهد آمد:

$$(X^TX+\alpha)W=X^T-\beta X^TXW1_k \quad (۱۲)$$

که در آن با استفاده از ماتریس W در مرحله tام، ماتریس Wدر مرحله $t+1$ ام محاسبه خواهد شد. لذا بر اساس عبارت (۱۲) رابطه تکراری زیر را خواهیم داشت:

$$(X^TX+\alpha I)W^{t+1}=X^TY-\beta X^TXW^t1_k \quad (۱۳)$$

در ادامه، با توجه به روابط به‌دست‌آمده، الگوریتم انتخاب ویژگی چند-برچسبه بر اساس مسئله رگرسیون کمترین مربعات، کمینه کردن نرخ افزونگی و تنک‌سازی تنک به‌صورت الگوریتم (۱) ارائه می‌شود.

الگوریتم ۱: انتخاب ویژگی چند-برچسبه بر اساس مسئله رگرسیون کمترین مربعات، کمینه کردن نرخ افزونگی و تنک‌سازی (FSRS).

ورودی: ماتریس داده $X\in\mathbb{R}^{n\times d}$, ماتریس برچسب $Y\in\mathbb{R}^{n\times d}$: پارامترهای α و β و تعداد ویژگی‌ها با تعداد k .

خروجی: زیرمجموعه I منتخب از ویژگی‌های با تعداد k.

t=0 :1

2: مقداردهی اولیه برای $W\in\mathbb{R}^{n\times d}$ به‌صورت تصادفی.

3: تکرار مراحل ٤ و ٥ تا زمان برقراری معیار همگرایی.

4: محاسبهٔ Wt^{+1} با استفاده از رابطه تکراری زیر:

$$(X^TX+\alpha I)W^{t+1}=X^TY-\beta X^TXW^t1_k$$

t=t+1 :٥

٦: مرتب‌سازی مقادیر $\|W_i^t\|$, که $i\in\{1,\dots,d\}$. به‌صورت نزولی و انتخاب مجموعه I به‌عنوان k عضو اول.

٦- نتایج عددی

به‌منظور بررسی کیفیت روش FSRS، عملکرد آن با برخی از روش‌های مشهور با استفاده از چهار مجموعه داده مشهور چند-برچسبه و در دسترس مقایسه خواهد شد. همچنین معیارهای ارزیابی، روش‌های استفاده‌شده برای مقایسه و نتایج حاصل به شرح زیر می‌باشند.

٦-١ مجموعه داده‌ها

به‌منظور انجام آزمایش‌های عددی، چهار مجموعه داده رایانه[1]، مرجع[2]، چشم‌انداز[3] و مخمر[4] که دارای کاربردهای متفاوت هستند، بررسی خواهند شد. داده‌های رایانه و مرجع به‌طور گسترده‌ای برای دسته‌بندی اتوماتیک صفحات وب بکار برده می‌شوند. داده چشم‌انداز، برای حل مسئله «به وجود آمدن تقاضا» در دسته‌بندی تصاویر معنایی، ایجادشده است. داده بیولوژیکی مخمر متعلق به مجموعه داده‌های میکرو-آرایه است که در آن هر ژن با استفاده از داده‌های بیان ریزآرایه الحاقی و پروفایل فیلوژنتیک توصیف‌شده است. لازم به ذکر است مجموعه داده‌های انتخاب‌شده فوق از

[1] Computer
[2] Reference
[3] Scene
[4] Yeast

کتابخانه مولان[1] قابل دریافت هستند و اطلاعات آن‌ها در جدول زیر جمع‌آوری‌شده است.

جدول (۱): مجموعه داده‌های استفاده‌شده

مجموعه داده	تعداد نمونه	تعداد ویژگی	تعداد برچسب
مرجع (Reference)	۵۰۰۰	۷۹۳	۳۳
رایانه (Computer)	۵۰۰۰	۶۸۱	۳۳
چشم‌انداز (Scene)	۲۴۰۷	۲۹۴	۶
مخمر (Yeast)	۲۴۱۷	۱۰۳	۱۴

۶-۲ معیارهای ارزیابی

ارزیابی عملکرد سیستم‌های چند-برچسبه یادگیری، متفاوت از سیستم‌های تک-برچسبه هستند. سیستم‌های یادگیری چند-برچسبه نیازمند معیارهای ارزیابی پیچیده‌تری هستند. فرض کنید یک مجموعه داده آزمون به‌صورت زیر باشد:

$$\tau = \{(x_i, y_i) | 1 \le i \le p\}$$

که در آن p تعداد نمونه‌های آزمون است. با توجه به نمونه x_i بردار برچسب باینری که توسط طبقه‌بندی چند-برچسبه پیش‌بینی‌شده است به‌عنوان $h(x_i)$ مشخص می‌شود، درحالی‌که رتبه پیش‌بینی‌شده برای k امین برچسب به‌عنوان $rank(x_i, k)$ تعریف‌شده است. معیارهای ارزیابی چندمنظوره به‌صورت زیر استفاده می‌شود:

۶-۲-۱ متوسط دقت

میانگین کسری از رتبه برچسب‌های بالاتر از یک برچسب خاص را در مجموعه برچسب‌های اساسی ارزیابی می‌کند.

$$AP(\tau) = \frac{1}{P}\sum_{i=1}^{p}\frac{1}{|y_i|}\sum_{y\in y_i}\frac{\left|\left\{y' | rank\left(x, y'\right) \le rank(x,y), y' \in y_i\right\}\right|}{rank(x_i, y)}$$

که مقدار $AP(\tau)$ هرچه قدر بیشتر باشد، کارایی بهتر خواهد بود.

۶-۲-۲ پوشش

ارزیابی تعداد گام‌هایی که به‌طور متوسط نیاز است تا لیست برچسب‌ها در برگرفته شوند تا همه برچسب‌های واقعی در نمونه را پوشش دهند.

$$CV(\tau) = \frac{1}{p}\sum_{i=1}^{p}\max_{y\in y_i} rannk(x_i, y) - 1$$

که مقدار $CV(\tau)$ ، هرچه قدر کوچک‌تر باشد، کارایی بهتر خواهد بود.

۶-۲-۳ افت همینگ

مقدار جفت برچسب‌های نامناسب طبقه‌بندی‌شده را محاسبه می‌کند. به‌عبارت‌دیگر، یک برچسب یا متعلق به نمونه‌های پیش‌بینی‌نشده است یا برچسبی است که متعلق به نمونه پیش‌بینی‌شده، نیست.

$$\mathrm{HL}(\tau) = \frac{1}{\mathrm{p}}\sum_{\mathrm{i}=1}^{\mathrm{p}}\frac{1}{\mathrm{q}}|\mathrm{h}(\mathrm{x_i})\Delta \mathrm{y_i}|$$

که در آن Δ نشان‌دهنده تفاضل متقارن بین دو مجموعه است. اینجا، مقدار $HL(\tau)$ هرچه قدر کوچک‌تر باشد، کارایی بهتر خواهد بود.

۶-۲-۴ یک - خطا

ارزیابی چندین‌باره برای بالاترین برچسبی که در مجموعه برچسب‌های مربوط به نمونه نیست.

$$OE(\tau) = \frac{1}{p}\sum_{i=1}^{p}\left\{\left[arg\max_{y\in Y} f(x_i, y)\right] \notin y_i\right\}$$

که مقدار OE هرچه قدر کوچک‌تر باشد، کارایی بهتر خواهد بود.

۶-۲-۵ افت رتبه

میانگین جفت برچسب‌هایی که برچسب نامناسب را بالاتر از برچسب مربوطه ارزیابی می‌کنند.

[1] Mulan Library: http://mulan.sourceforge.net/datasets-mlc.html.

$$RL(\tau) = \frac{1}{p}\sum_{i=1}^{p}\frac{1}{|y_i|\left|\overline{y}_l\right|}\left|\left\{\left(y', y''\right) | f\left(x_i, y'\right) \leq f\left(x_i, y''\right), \left(y', y''\right) \in y_i \times \overline{y}_i\right\}\right|$$

که $\overline{y}_i$ نشان‌دهنده متمم y_iاست. همچنین، مقدار $RL(\tau)$ هرچه قدر کوچک‌تر باشد، کارایی بهتر خواهد بود.

۶-۳ تنظیم پارامترها

تعدادی پارامتر درروش پیشنهادی FSRS وجود دارند که نیاز است این پارامترها تنظیم[1] شوند. برای مثال، مقادیر پارامترهای αوβ، به ترتیب، در مجموعه $\{10^{6},...,10^{1}\}$ و $\{10^{-1},...,10^{-6}\}$جستجو می‌شوند. در تمام روش‌ها، تعداد ویژگی‌های انتخاب‌شده k در مجموعه $\{10,20,...,100\}$ جستجو می‌شود. برای روش FSRS، بیشینه تعداد تکرار برابر با ۳۰ در نظر گرفته می‌شود. همچنین، روش k-fold به ازاء k=5، برای تنظیم کردن پارامترها در نظر گرفته‌شده است. به این معنی که داده را به ۵ قسمت تقسیم کرده و هربار یک‌پنجم داده به‌عنوان داده تست و مابقی به‌عنوان داده آموزش در نظر گرفته‌شده است.

۶-۴ آزمون‌های مقایسه

به‌منظور بررسی کیفیت روش پیشنهادی FSRS، عملکرد این روش با چند روش انتخاب ویژگی مؤثر چند-برچسبه مقایسه می‌شود. این روش‌ها عبارت هستند از:

روش PMU [۱۲]: یک روش انتخاب ویژگی چند-برچسبه مبتنی بر استفاده از اطلاعات مشترک و به حداکثر رساندن وابستگی بین ویژگی‌ها و برچسب‌ها.

روش MDMR [۱۰]: یک روش انتخاب ویژگی مبتنی بر ترکیب کردن «حداکثر وابستگی» و «کمترین افزونگی».

روش FIMF [۱۹]: یک روش انتخاب ویژگی چند- برچسبه سریع بر اساس رتبه‌بندی ویژگی‌های علمی اطلاعات.

۶-۵ نتایج تجربی

در این بخش، روش انتخاب ویژگی چند- برچسبه پیشنهادی FSRS، بر روی چهار مجموعه داده وسیع و در دسترس، با آزمون‌های گسترده مورد آزمایش قرارگرفته می‌شود، و نتایج حاصل با چندین الگوریتم انتخاب ویژگی کارا، که در بخش ۶ - ۴ معرفی شدند، مقایسه و عملکرد روش پیشنهادی ازنظر طبقه‌بندی موردبررسی قرار خواهد گرفت. برای این منظور، کارایی روش‌های انتخاب ویژگی چند-برچسبه با در نظر گرفتن معیارهای «متوسط دقت»، «پوشش»، «افت همینگ»، «یک -خطا» و «افت رتبه» ارزیابی می‌شوند و نتایج این بررسی‌ها در جدول‌های زیر به‌صورت کامل آورده شده است. در این جدول‌ها، در هر ردیف، بهترین نتیجه به‌صورت بولد و دومین بهترین نتیجه به‌صورت آندرلاین نشان داده می‌شوند.

جدول (۲): کارایی الگوریتم‌های مختلف برای معیار «متوسط دقت»

روش \ داده	FSRS	MDMR	PMU	FIMF
رایانه	**۰/۶۳۴۵**	۰/۶۳۰۵	۰/۶۰۹۲	۰/۶۲۰۳
مرجع	۰/۷۲۳۹	**۰/۷۲۵۷**	۰/۶۵۹۴	۰/۷۰۸۹
چشم‌انداز	**۰/۸۳۷۴**	۰/۷۶۳۳	۰/۸۰۳۴	۰/۶۹۰۶
مخمر	**۰/۷۵۹۰**	۰/۷۵۸۰	۰/۷۵۶۳	۰/۷۵۵۲

جدول (۳): کارایی الگوریتم‌های مختلف برای معیار «پوشش»

روش \ داده	FSRS	MDMR	PMU	FIMF
رایانه	٤/۳۸۱۳	**٤/۳٤۸۰**	٤/٦٥۲۳	٤/٥٥۸۰
مرجع	**۳/۰۹۱۳**	۳/۱٦٤۷	۳/٥٦۸۷	۳/۱۷۳
چشم‌انداز	**۰/٥۳۲٦**	۰/۸۲٥۳	۰/۷٤۹۲	۱/۰۹٥۳
مخمر	**٦/۲۹٦٦**	٦/۳٦٤۲	٦/۳۷۰۸	٦/۳۷٤۰

[1] Tune

جدول (٤): کارایی الگوریتم‌های مختلف برای معیار «افت همینگ»

روش \ داده	FSRS	MDMR	PMU	FIMF
رایانه	۰/۰۳۹۹	**۰/۰۳۹۸**	۰/۰٤۱٦	۰/۰٤۰۹
مرج	۰/۰۳۹٥	**۰/۰۳۹۱**	۰/۰٤٥۷	۰/۰٤۰۷
چشم‌انداز	**۰/۱۰۱٤**	۰/۱۳٤۸	۰/۱۱۳۷	۰/۱٥۸۷
مخمر	**۰/۱۹٥۹**	۰/۱۹۹۹	۰/۲۰۰٦	۰/۲۰۲۱

جدول (٥): کارایی الگوریتم‌های مختلف برای معیار «یک-خطا»

روش \ داده	FSRS	MDMR	PMU	FIMF
رایانه	**۰/٤٤۲۰**	۰/٤٥٤۳	۰/٤۷۰۰	۰/٤٦۲۷
مرجع	۰/۳٤۷۷	**۰/۳۳۷۳**	۰/۳۷۳۰	۰/۳۷۳۰
چشم‌انداز	**۰/۲۷٥۹**	۰/۳۰۹٤	۰/۳۰۹٤	۰/٤۹۸۳
مخمر	**۰/۲۲۳٦**	۰/۲۳٦٦	۰/۲۳۷۲	۰/۲۳۷٦

جدول (٦): کارایی الگوریتم‌های مختلف برای معیار «افت رتبه».

روش \ داده	FSRS	MDMR	PMU	FIMF
رایانه	**۰/۰۹۰۱**	۰/۰۹۰۳	۰/۰۹۸۰	۰/۰۹٥٥
مرجع	**۰/۰٥٥٤**	۰/۰٥٦٦	۰/۰٦۷۹	۰/۰٥۷۹
چشم‌انداز	**۰/۰۹۳٤**	۰/۱٤٤٤	۰/۱۲۹۰	۰/۱۹۹٤
مخمر	**۰/۱٦۹۲**	۰/۱۷۱۰	۰/۱۷۲۳	۰/۱۷٤۷

با توجه به معیارهای ارزیابی، بیشتر شدن عدد حاصل در معیار «متوسط دقت»، و در مابقی معیارها کمترشدن آن معیارها، بیانگر کارایی بهتر روش خواهد بود. با توجه به آنچه که در جدول‌های ۲ تا ٦ قابل‌مشاهده است، نتایج زیر حاصل می‌شوند:

(۱) در ارتباط با مجموعه داده‌های «چشم‌انداز» و «مخمر»، روش پیشنهادی FSRS قادر است در تمام معیارهای ارزیابی نسبت به سایر روش‌ها بهتر عمل کند.

(۲) در ارتباط با داده «مرجع» که روش پیشنهادی FSRS در رتبه بهترین دوم برحسب معیارهای ارزیابی «افت همینگ» و «یک -خطا» قرار دارد، روش پیشنهادی قادر است در تمام معیارهای ارزیابی دیگر نسبت به سایر روش‌ها بهتر عمل کند.

(۳) در ارتباط با داده «رایانه» که روش پیشنهادی FSRS در رتبه بهترین دوم برحسب معیارهای ارزیابی «پوشش» و «افت همینگ» قرار دارد، روش پیشنهادی قادر است در تمام معیارهای ارزیابی دیگر نسبت به سایر روش‌ها بهتر عمل کند.

٦-٦ آزمون‌های آماری غیر پارامتری

در این بخش، به‌منظور دستیابی به یک توجیه منصفانه آماری در بررسی تفاوت‌ها و برتری روش پیشنهادی در مقایسه با دیگر روش‌های اشاره‌شده، از یک آزمون آماری مانند آزمون فریدمن، استفاده می‌شود. آزمون فریدمن یک آزمون آماری ناپارامتری است که از آن برای مقایسه میانگین رتبه‌ها در بین k متغیر (گروه) استفاده می‌شود. برای این منظور، یک آزمون فرض به‌صورت مقدار احتمال (p-مقدار) برابر با کمترین مقدار از سطح معنی‌دار، که موجب رد فرض صفر خواهد شد، در نظر گرفته می‌شود. جدول ۷ نتایج حاصل از آزمون فریدمن را برای هر روش و برحسب هریک از معیارهای ارزیابی نشان می‌دهد. در اینجا، هرچه رتبه روش پایین‌تر باشد، آنگاه کارایی آن روش بهتر است. مطابق جدول ۷، مشاهده می‌شود که روش FSRS پایین‌ترین رتبه را برحسب هریک از معیارهای ارزیابی دارد، و بنابراین می‌توان نتیجه گرفت که روش FSRS در مقایسه با سایر روش‌ها از کارایی بالاتری برخوردار است.

جدول (۷): میانگین رتبه روش‌ها حاصل از آزمون آماری فریدمن و بر اساس معیارهای مختلف

(آ) میانگین دقت

روش	رتبه
FSRS	**۱/۲٥**
MDMR	۲
PMU	۳/۲٥

FIMF	۳/۵

(ب) پوشش

روش	رتبه
FSRS	**۱/۲۵**
MDMR	۲
PMU	۳/۲۵
FIMF	۳/۵

(ج) افت همینگ

روش	رتبه
FSRS	**۱/۵**
MDMR	۱/۷۵
PMU	۳/۲۵
FIMF	۳/۵

(د) یک-خطا

روش	رتبه
FSRS	**۱/۲۵**
MDMR	۲/۱۲۵
PMU	۳/۲۵
FIMF	۳/۳۷۵

(ه) افت رتبه

روش	رتبه
FSRS	**۱**
MDMR	۲/۲۵
PMU	۳/۲۵
FIMF	۳/۵

به‌منظور تحقیق بیشتر راجع به کیفیت نتایج آماری، مقایسه اندازه p-مقدار حاصل از آزمون آماری فریدمن بین روش FSRS و سایر روش‌ها در جدول ۸ نمایش داده‌شده است. در اینجا، سطح معنی‌دار αبه‌صورت $\alpha = ۰/۰۱۶۶۶۷$ در نظر گرفته‌شده است. همچنین، توجه شود که آزمون آماری فریدمن، روش‌هایی با اندازه p-مقدار کوچک‌تر یا مساوی از ۰/۰۱۶۶۶۷ را رد می‌کند.

جدول (۸): مقایسه اندازهp - مقدار حاصل از آزمون آماری فریدمن بین روش FSRS و سایر روش‌ها.

(آ) میانگین دقت

رتبه	روش	p - مقدار
۳	FIMF	۰/۰۱۳۷۱۱
۲	PMU	۰/۰۲۸۴۶
۱	MDMR	۰/۴۱۱۳۱۴

(ب) پوشش

رتبه	روش	p - مقدار
۳	FIMF	۰/۰۱۳۷۱۱
۲	PMU	۰/۰۲۸۴۶
۱	MDMR	۰/۴۱۱۳۱۴

(ج) افت همینگ

رتبه	روش	p - مقدار
۳	FIMF	۰/۰۲۸۴۶
۲	PMU	۰/۰۵۵۲۳۴
۱	MDMR	۰/۷۸۴۱۹۱

(د) یک-خطا

رتبه	روش	p - مقدار
۳	FIMF	۰/۰۱۹۹۲۲
۲	PMU	۰/۰۲۸۴۶
۱	MDMR	۰/۳۳۷۸۰۳

(ه) افت رتبه

رتبه	روش	p - مقدار
۳	FIMF	۰/۰۰۶۱۷
۲	PMU	۰/۰۱۳۷۱۱
۱	MDMR	۰/۱۷۰۹۰۴

توجه شود که در این آزمایش، روش که اندازهp - مقدار آن کوچک‌تر یا مساوی مقدار ۰/۰۱۶۶۶۷ باشد، رد می‌شود.

با توجه به جدول ۸، می‌توان نتیجه گرفت که روش پیشنهادی FSRS روش FIMF را در معیارهای میانگین دقت، پوشش و افت رتبه به‌طور کامل رد می‌کند. همچنین، روش PMU فقط در معیار افت رتبه رد شده است؛ اما مقادیر p–مقدار در سایر روش‌ها نزدیک به مقدار α است که نشان‌دهنده این است که این روش فاصله زیادی تا رد شدن توسط روش پیشنهادی ما ندارد.

۷- نتیجه‌گیری

در این مقاله، یک روش انتخاب ویژگی چند–برچسبه مبتنی بر «رگرسیون کمترین مربعات»، «کمینه کردن افزونگی» و «تنظیم تنکی»، به نام روش FSRF معرفی شد، که در این روش با استفاده از مسئله کمینه‌سازی کمترین مربعات خطا، مسئله انتخاب ویژگی به‌صورت یک مسئله تجزیه ماتریسی همراه با جمله منظم‌سازی تنک و با استفاده از نرم فروبنیوس بیان می‌شود. برای این منظور، یک مسئله بهینه‌سازی برای روش FSRS طراحی و یک الگوریتم بروز رسانی تکراری نیز برای حل تابع هدف متناظر با مسئله انتخاب ویژگی مطرح‌شده، موردبررسی قرار گرفت. مجموعه‌ای از آزمایش‌های عددی جهت بررسی کارایی روش پیشنهادی FSRS و مقایسه عملکرد این روش با عملکرد چند روش انتخاب ویژگی چند–برچسبه کارآمد انجام گرفت. نتایج این آزمایش‌ها نشان دادند که روش پیشنهادشده دارای عملکرد مطلوب و رضایت‌بخشی است.

References

[1] McCallum, A. (1999). Multi-Label Text Classification with a Mixture Model Trained by EM. AAAI'99 Workshop on Text Learning.

[2] Boutell, M. R., Luo, J., Shen, X. & Brown, C. M. (2004). Learning Multi-Label Scene Classification. Pattern Recognition, 37, 1757--1771.

[3] Zhang, M.-L. & Zhou, Z.-H. (2006). Multi-Label Neural Networks with Applications to Functional Genomics and Text Categorization.. IEEE Trans. Knowl. Data Eng., 18, 1338-1351.

[4] Xu, X., Shan, D., Li, S., Sun, T., Xiao, P., & Fan, J. (2019). Multi-label learning method based on ML-RBF and laplacian ELM. Neurocomputing, 331, 213–219.

[5] N. Zhang, S. Ding, J. Zhang, Multi layer ELM-RBF for multi-label learning, Applied SoftComputing, 43, 535-545, 2016.

[6] Wang, S., Pedrycz, W., Zhu, Q. & Zhu, W. (2015). Subspace learning for unsupervised feature selection via matrix factorization.. Pattern Recognit., 48, 10-19.

[7] Wang, S., Pedrycz, W., Zhu, Q. & Zhu, W. (2015). Unsupervised feature selection via maximum projection and minimum redundancy.. Knowl. Based Syst., 75, 19-29.

[8] Jian, L., Li, J., Shu, K., & Liu, H. (2016). Multi-label informed feature selection. IJCAI International Joint Conference on Artificial Intelligence, 2016-January, 1627-1633.

[9] Huang, R., & Wu, Z. (2021). Multi-label feature selection via manifold regularization and dependence maximization. Pattern Recognition, 120, 108149. DOI:10.1016/j.patcog.2021.108149

[10] Lin, Y., Hu, Q., Liu, J., & Duan, J. (2015). Multi-label feature selection based on max-dependency and min-redundancy. Neurocomputing, 168, 92–103. DOI:10.1016/j.neucom.2015.06.010

[11] Shannon, C. E. (2001). A Mathematical Theory of Communication. SIGMOBILE Mob. Comput. Commun. Rev., 5(1), 3–55. DOI:10.1145/584091.584093

[12] Lee, J., & Kim, D.-W. (2013). Feature selection for multi-label classification using multivariate mutual information. Pattern Recognition Letters, 34(3), 349–357. DOI:10.1016/j.patrec.2012.10.005

[13] Lee, J., & Kim, D.-W. (2015). Mutual Information-based multi-label feature selection using interaction information. Expert Systems with Applications, 42(4), 2013–2025. DOI:10.1016/j.eswa.2014.09.063.

[14] Lee, J., & Kim, D.-W. (2015). Memetic feature selection algorithm for multi-label classification. Information Sciences, 293, 80–96. DOI:10.1016/j.ins.2014.09.020

[15] Doquire, G., & Verleysen, M. (2013). Mutual information-based feature selection for

multilabel classification. Neurocomputing, 122, 148–155. DOI:10.1016/j.neucom.2013.06.035

[16] Jian, L., Li, J., Shu, K., & Liu, H. (2016). Multi-Label Informed Feature Selection. Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence, 1627–1633. New York, New York, USA: AAAI Press.

[17] Huang, R., Jiang, W., & Sun, G. (2018). Manifold-based constraint Laplacian score for multi-label feature selection. Pattern Recognition Letters, 112, 346–352. DOI:10.1016/j.patrec.2018.08.021

[18] Hu, J., Li, Y., Gao, W., & Zhang, P. (2020). Robust multi-label feature selection with dual-graph regularization. Knowledge-Based Systems, 203, 106126. DOI:10.1016/j.knosys.2020.106126

[19] Lee, J., & Kim, D.-W. (2015). Fast multi-label feature selection based on information-theoretic feature ranking. Pattern Recognition, 48(9), 2761–2771. DOI:10.1016/j.patcog.2015.04.009

Journal of Applied And Basic Machine Intelligence Research (ABMIR)

Applied and Basic Machine Intelligence Research

Vol. 1, No. 2, Autumn & Winter 2023, pp. 16 - 28

Multi-label feature selection method based on redundancy minimization

Hakimeh Hasaniyeh[1], Behzad Ebrahimi[2], Farid Saberi-Movahed[1], Mahdi Eftekhari[2*]

[1]Department of Applied Mathematics, Faculty of Sciences and Modern Technologies, Graduate University of Advanced Technology, Kerman, Iran

[2]Department of Computer Engineering, Faculty of Engineering, Shahid Bahonar University of Kerman, Kerman, Iran

Article Information

Original Research Paper

Received:
2022 September 20

Accepted:
2023 June 13

Keywords:
Multi-label feature selection, Redundancy minimization, Regression, Least squared error, Sparsity regularization

Corresponding Author*:
m.eftekhari@uk.ac.ir

Abstract

Feature selection methods are known to be effective in improving the learning process. The purpose of a feature selection method is to identify relevant features and remove irrelevant features in order to obtain a suitable subset of features, so that the redundancy between the selected features is minimized. In multi-label data, if there is a correlation between features, it is possible that the amount of redundancy in the feature set is increased. The existence of redundancy between features along with the challenge of high dimensions of multi-label data can grow the computational calculations, decrease the accuracy and finally increase the probability of errors in the prediction and classification of multi-label data. In this article, with the aim of minimizing the redundancy of features, a multi-label feature selection algorithm is proposed considering the least squares regression model and sparse regularization. Finally, using a number of well-known multi-label data sets, the efficiency of the proposed method is verified and the results are compared with some common multi-label feature selection methods.

: 10.22034/ABMIR.2023.18967.1016

E-ISSN: 2821-2037 /© 2023. Published by Yazd University This is an open access article under the CC BY 4.0 License (https://creativecommons.org/licenses/by/4.0/).