Journal of Applied And Basic Machine Intelligence Research (ABMIR)

پژوهش‌های نظری و کاربردی هوش ماشینی

سال سوم، شماره اول، بهار و تابستان ۱٤۰٤، صفحات۱۷۷ - ۱٦۳

انتخاب ویژگی نیمه‌نظارتی مبتنی‌بر خودرمزنگار گراف با حفظ ساختار محلی-گسترده

محمدجواد رضایی[۱]، مهدی‌آقا صرام[۲]، راضیه شیخ‌پور[۳]*

[۱]دانشجوی دکتری، دانشکده مهندسی کامپیوتر، دانشگاه یزد، یزد، ایران

[۲]دانشیار، دانشکده مهندسی کامپیوتر، دانشگاه یزد، یزد، ایران

[۳]دانشیار، گروه مهندسی کامپیوتر، دانشکده فنی و مهندسی، دانشگاه اردکان، اردکان، ایران

چکیده

پردازش داده‌های با ابعاد بالا چالش مهمی در حوزه‌های مختلف است و انتخاب ویژگی به‌عنوان روشی مؤثر برای کاهش ابعاد، نقش کلیدی در بهبود عملکرد مدل‌های یادگیری ماشین دارد. از آنجا که برچسب‌گذاری داده‌ها پرهزینه و زمان‌بر است، انتخاب ویژگی نیمه‌نظارتی که از داده‌های بدون برچسب نیز استفاده کند، اهمیت ویژه‌ای دارد. در این مقاله، یک روش انتخاب ویژگی نیمه‌نظارتی تنک مبتنی بر خودرمزنگار گراف ارائه می‌شود که دو نوآوری اصلی دارد: (۱) ترکیب خودرمزنگار برای حفظ ساختار کلی داده و گراف طیفی نیمه‌نظارتی برای حفظ ساختار محلی و اطلاعات برچسب (۲) اعمال منظم‌سازی نرم-$L_{(2,1)}$ برروی ماتریس وزن رمزگذار تا سطرهای غیرمؤثر به صفر میل کرده و ویژگی‌های نامرتبط به‌طور خودکار حذف شوند. بهینه‌سازی مسئله با الگوریتم گرادیان و پس‌انتشار انجام شده و مشتق منظم‌سازی در به‌روزرسانی پارامترها لحاظ می‌شود؛ بدین ترتیب انتخاب ویژگی به صورت درون‌مدلی و هم‌زمان با آموزش شبکه انجام می‌گیرد. روش پیشنهادی بر روی شش مجموعه‌داده استاندارد UCI شامل ORL، ATT، WBCD، WDBC، QSAR و پارکینسون ارزیابی و با پنج روش مرجع مقایسه شد. معیار ارزیابی، دقت طبقه‌بندی با استفاده از ماشین بردار پشتیبان و-k نزدیک‌ترین همسایه بود. نتایج دو طبقه‌بند برروی شش مجموعه داده به ترتیب ۰/۷۸، ۰/۸۸، ۰/۹۸، ۰/۹۷، ۰/۸۱، ۰/۹۱ و ۰/۷۵، ۰/۹۲، ۰/۹۷، ۰/۹٤، ۰/۸۲، ۰/۹۲ نشان داد که روش پیشنهادی در اغلب موارد عملکرد برتری دارد. این یافته‌ها تأیید می‌کنند که چارچوب پیشنهادی با بهره‌گیری همزمان از ساختار داده و منظم‌سازی تنک، قادر به انتخاب مجموعه‌ای کارآمد از ویژگی‌ها در شرایط نیمه‌نظارتی است.

مقاله پژوهشی

تاریخ دریافت:
۱٤۰٤/٤/۱۸
تاریخ پذیرش:
۱٤۰٤/۷/۲۰

کلیدواژه‌ها:
انتخاب ویژگی نیمه‌نظارتی، خودرمزنگار، مدل‌های تنک، منظم‌سازی نرم$L_{(2,1)}$

نویسنده‌مسئول:
rsheikhpour@ardakan.ac.ir

doi : 10.22034/ABMIR.2025.23363.1140

E-ISSN: 2821-2037

۱- مقدمه

داده‌ها توسط مجموعه‌ای از ویژگی‌ها توصیف می‌شوند که تعداد آن‌ها، ابعاد داده نامیده می‌شود. الگوریتم‌های یادگیری ماشین با بهره‌گیری از این ویژگی‌ها، دانش نهفته در داده‌ها را استخراج می‌کنند. در دهه‌های اخیر، افزایش چشمگیر حجم داده‌های با ابعاد بالا در حوزه‌های گوناگون از جمله بینایی ماشین، تشخیص الگو و داده‌کاوی، ذخیره‌سازی و پردازش این نوع داده‌ها را به یک چالش جدی تبدیل کرده است [۱]. اگرچه ابعاد بالای داده‌ها امکان ذخیره‌سازی اطلاعات بیشتر را فراهم می‌آورد، اما یادگیری از این داده‌ها اغلب با مشکلاتی همراه است. از جمله این مشکلات می‌توان به کاهش قابلیت تعمیم مدل‌های یادگیری ناشی از پدیده بیش‌برازش[1] در طی فرآیند آموزش اشاره کرد. علاوه بر این، تمامی ویژگی‌های ذخیره‌شده لزوماً برای استخراج دانش مفید نیستند و وجود ویژگی‌های نامربوط یا زائد، هم سربار محاسباتی را افزایش می‌دهد و هم احتمال بروز خطا را بالا می‌برد. این مسائل، چالش‌های مهمی را در تحلیل داده‌های با ابعاد بالا ایجاد می‌کنند [۲]. با این حال، تحقیقات نشان داده‌اند که ابعاد ذاتی بسیاری از داده‌های با ابعاد بالا، به مراتب کمتر از ابعاد ظاهری آن‌هاست. از این رو، کاهش ابعاد به عنوان یک گام پیش‌پردازش رایج و مؤثر در تحلیل این نوع داده‌ها مطرح می‌شود که هدف آن، کاهش زمان پردازش و بهبود قابلیت تعمیم مدل‌های یادگیری است [۳]. انتخاب ویژگی یکی از روش‌های پرکاربرد کاهش ابعاد است که هدف آن، انتخاب بهینه زیرمجموعه‌ای از ویژگی‌هاست. به‌طورکلی، روش‌های انتخاب ویژگی به سه دسته اصلی تقسیم می‌شوند: فیلتر[2]، بسته‌بندی[3] و جاسازی‌شده[4]. روش‌های فیلتر، مستقل از الگوریتم یادگیری، بر ویژگی‌های ذاتی مجموعه داده تکیه دارند و فرآیند انتخاب ویژگی را به عنوان یک گام پیش‌پردازش انجام می‌دهند. مزیت این روش‌ها، هزینه محاسباتی پایین و قابلیت تعمیم خوب آن‌هاست. روش‌های بسته‌بندی، از یک الگوریتم یادگیری برای ارزیابی میزان مفید بودن زیرمجموعه‌های مختلف ویژگی‌ها استفاده می‌کنند. معیار انتخاب در این روش‌ها، عملکرد الگوریتم یادگیری با استفاده از هر زیرمجموعه از ویژگی‌هاست. در مقایسه با روش‌های فیلتر، روش‌های بسته‌بندی معمولاً به عملکرد بهتری دست می‌یابند، اما هزینه محاسباتی بالاتری دارند. در روش‌های جاسازی‌شده، انتخاب ویژگی به صورت یکپارچه در فرآیند آموزش الگوریتم انجام می‌شود. به عبارت دیگر، جستجو برای زیرمجموعه بهینه ویژگی‌ها در حین ساخت مدل و با استفاده از یک تابع هدف صورت می‌گیرد [۴]. از منظر دیگر، روش‌های انتخاب ویژگی بر اساس نحوه استفاده از برچسب‌های داده به سه دسته باناظر، بدوناظر و نیمه‌نظارتی تقسیم می‌شوند [۵]. در روش‌های باناظر، ویژگی‌ها بر اساس توانایی آن‌ها در حفظ اطلاعات مربوط به برچسب کلاس ارزیابی می‌شوند [۶]. امروزه، حجم عظیمی از داده‌ها در کاربردهای دنیای واقعی تولید می‌شوند که تنها بخش کوچکی از آن‌ها دارای برچسب هستند. برچسب‌گذاری داده‌ها اغلب نیازمند صرف زمان و هزینه زیادی توسط متخصصان است. بنابراین، مجموعه نمونه‌های برچسب‌گذاری‌شده معمولاً برای آموزش روش‌های انتخاب ویژگی با ناظر کافی نیستند [۷]. در روش‌های بدون ناظر، هیچ اطلاعاتی در مورد برچسب کلاس در دسترس نیست و معیار ارزیابی ویژگی‌ها، حفظ ساختار محلی و یا ذاتی داده‌ها است [۸]. روش‌های بدون ناظر، اطلاعات ارزشمند موجود در داده‌های برچسب‌دار را نادیده می‌گیرند. یکی از روش‌های انتخاب ویژگی بدون ناظر خودرمزنگار است که با حفظ ساختار داده گسترده و حفظ ساختار ذاتی داده‌ها سعی در انتخاب یک زیرمجموعه مناسب از ویژگی‌ها دارد. روش‌های نیمه‌نظارتی، با بهره‌گیری از اطلاعات موجود در هر دو دسته داده‌های برچسب‌دار و بدون برچسب، به دنبال انتخاب ویژگی‌های مناسب هستند که هم اطلاعات برچسب کلاس داده‌های برچسب دار و هم ساختار محلی داده‌های برچسب دار و بدون برچسب را در نظر می‌گیرند [۹].

با وجود پیشرفت‌های اخیر، چالش‌های متعددی باقی مانده است. بسیاری از روش‌های نیمه‌نظارتی موجود تنها یکی از جنبه‌های

[1] Overfitting
[2] Filter
[3] Wrapper
[4] Embedded

ساختار داده (مثلاً محلی یا کلی) را در نظر می‌گیرند و بنابراین تصویر ناقصی از توزیع داده ارائه می‌دهند. برخی دیگر به دلیل اعمال محدودیت‌های سخت‌گیرانه روی ماتریس تبدیل، انعطاف‌پذیری کافی برای یافتن زیرمجموعه‌ای بهینه از ویژگی‌ها ندارند. علاوه بر این، اکثر کارهای پیشین کمتر به ارتباط مستقیم بین فرایند بازنمایی داده و انتخاب ویژگی توجه داشته‌اند.

در این مقاله، یک چارچوب نوین برای انتخاب ویژگی نیمه‌نظارتی تنک مبتنی بر خودرمزنگار گراف ارائه شده است که همزمان سه نوع اطلاعات را در نظر می‌گیرد: (۱) ساختار کلی داده از طریق بازسازی خودرمزنگار، (۲) ساختار محلی و برچسب‌ها با استفاده از گراف طیفی نیمه‌نظارتی، و (۳) تنکی سطری ماتریس وزن رمزگذار برای حذف خودکار ویژگی‌های غیرمؤثربا استفاده از عبارت منظم‌سازی نرم-$L_{2,1}$. این ترکیب باعث می‌شود ویژگی‌های منتخب هم از نظر آماری متمایز و هم با ساختارهای واقعی داده سازگار باشند. روش پیشنهادی روی شش مجموعه‌داده استاندارد UCI شامل ORL، ATT، WBCD، WDBC، QSAR و پارکینسون آزمایش شده و با پنج روش مرجع مقایسه گردید. نتایج نشان می‌دهد که روش پیشنهادی در اکثر موارد برتری محسوسی دارد؛ به‌طور نمونه در مجموعه‌داده ORL دقت طبقه‌بندی ۷۸٪ به دست آمد که نسبت به بهترین روش مرجع بیش از ۱۲ درصد بهبود را نشان می‌دهد.

سازماندهی ادامه مقاله بدین شرح است. در بخش دوم، کارهای مرتبط مرور می‌شود. در بخش سوم، به تشریح روش پیشنهادی، نحوه بهینه‌سازی و الگوریتم تکراری حل آن پرداخته می‌شود. بخش چهارم، به نحوه ارزیابی روش پیشنهادی و همچنین مطالعه همگرایی الگوریتم پیشنهادی می‌پردازد. در بخش پنجم، نتیجه‌گیری مقاله ارائه می‌شود.

۲- کارهای مرتبط

۲-۱ روش‌های انتخاب ویژگی نیمه‌نظارتی تنک

در سال‌های اخیر، روش‌های انتخاب ویژگی تنک[1] به سرعت توسعه یافته‌اند. هدف این روش‌ها، استفاده از مدل‌های تنک به منظور انتخاب ویژگی‌هایی است که علاوه بر تنک بودن، دارای بیشترین قدرت تفکیک‌پذیری نیز باشند [۱۰]. نرم-L_1[2] به‌عنوان یکی از شناخته‌شده‌ترین مدل‌های تنک، کاربرد فراوانی دارد. با این وجود، روش‌های انتخاب ویژگی پراکنده مبتنی‌بر نرم-L_1، گاهی اوقات قادر به انتخاب ویژگی‌هایی با میزان تنک‌شدگی کافی نیستند. در راستای بهبود این محدودیت، پژوهش‌های متعددی، تعمیم نرم-L_1 به نرم-L_p $(0<p<1)$ را موردبررسی قرار داده‌اند تا نمایش تنک‌تری از داده‌ها حاصل شود. به عنوان مثال، ژو و همکاران نشان دادند که مدل نرم-$L_{1/2}$ بهترین عملکرد و پراکندگی را ارائه می‌دهد [۱۱]. با این حال، مدل‌های مذکور، اطلاعات مربوط به همبستگی بین ویژگی‌های مختلف را در نظر نمی‌گیرند. تحقیقات اخیر، مزایای در نظر گرفتن همبستگی بین ویژگی‌ها در فرآیند انتخاب ویژگی را نشان داده‌اند. نی و همکاران با معرفی منظم‌سازی نرم-$L_{2,1}$ چارچوبی کارا برای انتخاب ویژگی مقاوم ارائه دادند [۱۲]. ونگ و چن با تعمیم به نرم-$L_{2,p}$ مدل با انعطاف‌پذیری بیشتری در کنترل تنکی به دست آوردند [۱۳]. این روش‌ها همبستگی بین ویژگی‌ها را در انتخاب ویژگی لحاظ می‌کند. شی و همکاران انتخاب ویژگی مبتنی بر گراف لاپلاسین را برای داده‌های تصویری پیشنهاد کردند[۱۰] به دنبال آن شیخ‌پور و همکاران روشی مقاوم مبتنی بر گراف برای انتخاب ویژگی تنک نیمه‌نظارتی ارائه کردند [۱۴]. لی و همکاران در چارچوبی نیمه‌نظارتی، انتخاب ویژگی را با ترکیب محدودیت‌های هندسی و اطلاعات برچسب‌ها انجام دادند [۱۵]. در زمینه انتخاب ویژگی نیمه‌نظارتی مبتنی‌بر مدل‌های تنک، هدف، محاسبه یک ماتریس انتقال $W \in R^{d\times c}$ است که به صورت بهینه، اطلاعات مربوط به تفکیک‌پذیری و توزیع داده‌های آموزشی را حفظ کند. هر سطر از ماتریس W برای وزن‌دهی به ویژگی‌ها استفاده می‌شود. در صورتی که برخی از سطرهای W صفر باشند، می‌توان از W برای انتخاب ویژگی استفاده کرد؛ به این ترتیب که ویژگی‌های مرتبط با سطرهای غیر صفر در W انتخاب می‌شوند. برای حصول اطمینان از تنک

[1] Sparse Feature Selection

[2] Lasso

بودن سطرهای ماتریس W و همچنین برای جلوگیری از هموار بودن تابع هدف، یک عبارت منظم‌سازی مبتنی بر نرم- L_1 یا نرم- L_p به تابع هدف اضافه می‌شود.

با وجود این پیشرفت‌ها، اکثر این روش‌ها یا تنها از ساختار محلی داده استفاده می‌کنند یا تنها به تنکی توجه دارند و کمتر به ترکیب هم‌زمان چندین منبع اطلاعاتی (ساختار کلی، ساختار محلی و برچسب‌ها) پرداخته‌اند.

۲-۲ خودرمزنگار

خودرمزنگار[1]ها برای اولین بار توسط رملهارت و همکاران به‌عنوان نوع خاصی از شبکه‌های عصبی معرفی شدند که هدف اصلی آن‌ها یادگیری بازنمایی‌های فشرده و معنادار از داده‌ها به صورت بدون‌نظارت است [۱۶]. این شبکه‌ها از دو بخش اصلی تشکیل شده‌اند: یک رمزگذار۲ که داده‌های ورودی ($x \in R^d$) را به یک فضای با ابعاد کمتر ($y \in R^m$) نگاشت می‌کند (d > m)، و یک رمزگشا۳ تلاش می‌کند این نمایش فشرده را به فضای ورودی اصلی بازگرداند و داده‌ها را بازسازی کند [۱۷]. به عبارت دیگر، برخلاف شبکه‌های عصبی سنتی که برای پیش‌بینی مقدار هدف Y بر اساس ورودی X آموزش داده می‌شوند، خودرمزنگارها برای بازسازی ورودی خود (X) آموزش می‌بینند.

هدف اصلی در آموزش یک خودرمزنگار، به حداقل رساندن تفاوت بین داده‌های ورودی (x) و داده‌های بازسازی‌شده (x') است. این تفاوت معمولاً با استفاده از یک تابع هزینه ٤ اندازه‌گیری می‌شود.

ویژگی مهم خودرمزنگار برای انتخاب ویژگی این است که ماتریس وزن لایه رمزگذار، بیانگر میزان مشارکت هر ویژگی ورودی در بازنمایی فشرده است؛ بنابراین با اعمال منظم‌سازی مناسب روی این وزن‌ها می‌توان ویژگی‌های مؤثر را شناسایی کرد. پژوهش‌های اخیر این ایده را در چارچوب انتخاب ویژگی توسعه داده‌اند. فنگ و دوارت با ترکیب خودرمزنگار و گراف، روشی برای انتخاب ویژگی بی‌نظارت ارائه دادند [۱۸]. شنگ و همکاران نیز از بازنمایی خودبیان‌گر همراه با تنظیم گراف برای انتخاب ویژگی استفاده کردند [۱۹]. در سال‌های اخیر، خودرمزگذارها به‌عنوان ابزاری قدرتمند برای یادگیری بازنمایی در حوزه نیمه‌نظارتی نیز استفاده شده است. وینسنت و همکاران که از پیشگامان این حوزه بودند با معرفی خودرمزگذار انباشته کاهنده نویز نشان دادند که می‌توان از پیش‌آموزش بدون‌ناظر برای استخراج بازنمایی‌های سطح بالا از داده‌های بدون برچسب استفاده کرد و سپس این بازنمایی‌ها را در فرآیند تنظیم دقیق طبقه‌بند با داده‌های برچسب‌دار محدود به‌کار گرفتند که منجر به بهبود چشمگیر عملکرد طبقه‌بندی شدند [۲۰]. در ادامه، کینگما و ولینگ خودرمزگذار واریشنال (VAE) را به‌عنوان یک چارچوب مولد معرفی کردند که با مدل‌سازی یک فضای نهان مشترک، قابلیت استفاده مؤثر از داده‌های برچسب‌دار و بدون‌برچسب را فراهم می‌سازد [۲۱]. یکی از رویکردهای نوآورانه در یادگیری نیمه‌نظارتی، استفاده از خودرمزگذار متخاصم است. این چارچوب نخستین بار توسط مخزنی و همکاران معرفی شد و به‌عنوان یک مدل مولد عمل می‌کند که از یک شبکه متخاصم برای واداشتن توزیع فضای نهان خودرمزگذار به تبعیت از یک توزیع پیشین دلخواه (مانند توزیع گاوسی) بهره می‌گیرد. این مکانیزم موجب می‌شود فضای نهان آموخته‌شده ساختاریافته، منظم و قابل‌تفسیر باشد و درنتیجه برای وظایف طبقه‌بندی و تحلیل داده‌های نیمه‌نظارتی کارایی بالاتری داشته باشد [۲۲].

با وجود پیشرفت‌های چشمگیر در استفاده از خودرمزنگار و روش‌های گراف‌محور در یادگیری نیمه‌نظارتی، اکثر این رویکردها همچنان با محدودیت‌هایی همراه هستند. بیشتر این رویکردها یا در حالت بی‌نظارت عمل می‌کنند و از اطلاعات برچسب صرف‌نظر می‌کنند، یا تنها بر یکی از ساختارهای داده (مثلاً محلی یا کلی) تمرکز دارند.

[1] Autoencoder
[2] Encoder
[3] Decoder
[4] Loss Function

جدول (۱): مقایسه روش پیشنهادی با سایر روش‌های انتخاب ویژگی رایج

روش	استفاده از خودرمزنگار	حفظ ساختار گسترده	حفظ ساختار محلی	استفاده از داده‌های برچسب‌دار	محدودیت متعامد	نوع منظم‌سازی	توضیح متمایزکننده
Vincent et al.	✓	✓	✗	✗	✗	✗	پیش‌آموزش بدون‌ناظر برای یادگیری بازنمایی‌های سطح بالا برای تنظیم دقیق طبقه‌بند؛ فاقد درنظر گرفتن داده‌های برچسب‌دار و ساختار محلی
Kingma & Welling	✓	✓	✗	به صورت غیرمستقیم	احتمالی	✗	چارچوب مولد برای مدل‌سازی فضای نهان؛ قابلیت استفاده در نیمه‌نظارتی ولی بدون درنظر گرفتن ساختار محلی
Makhzani et al	✓	✓	✗	به صورت نیمه‌نظارتی	متکی بر توزیع پیشین (مانند گاوسی)	✗	ترکیب AE با شبکه متخاصم برای منظم‌سازی فضای نهان؛ یادگیری فضای فشرده منظم و قابل‌تفسیر برای نیمه‌نظارتی
Feng et al.	✓	✓	✗	✗	✓	✗	فقط ساختار گسترده داده را باAE حفظ می‌کند؛ فاقد اطلاعات برچسبی و ساختار محلی
Sheikhpour et al.	✗	✗	✓	✓	✓	$L_{2,1}$	مبتنی بر گراف و منظم‌سازی؛ فاقد بازسازی داده باAE
Li et al	✗	✗	✓	✓	✓	محدودیت عدم همبستگی (uncorrelated constraint)	بر گراف و قیدهای هندسی تکیه دارد؛ فاقد مکانیزم بازسازی یا تنکیAE
Wang et al.	✗	✗	✓	✓	✓	Structured sparse	انتخاب ویژگی تبعیضی با ساختار تنک؛ اما فاقد خودرمزنگار و حفظ ساختار گسترده
سایر روش‌های نیمه‌نظارتی	✗	وابسته به روش	✓	✓	اغلب	متنوع	معمولاً تنها یکی از جنبه‌های داده (محلی یا برچسبی) را لحاظ می‌کنند
روش پیشنهادی (GASFS)	✓	✓	✓	✓	✗	$L_{2,1}$	ترکیب نوآورانه خودرمزنگار، گراف نیمه‌نظارتی، منظم‌سازی $L_{2,1}$ بدون محدودیت متعامد بنابراین انعطاف‌پذیری و انتخاب ویژگی کارآمدتر

شکاف اصلی در ادبیات این است که تاکنون چارچوبی که بتواند به‌طور هم‌زمان بازسازی کلی داده‌ها، حفظ ساختار محلی و بهره‌گیری از برچسب‌ها را با منظم‌سازی تنک ادغام کند، کمتر مورد توجه قرار گرفته است. بنابراین نیاز به روشی وجود دارد که علاوه بر استفاده از مزایای خودرمزنگارها در یادگیری بازنمایی کلی، ساختار محلی داده‌ها و اطلاعات برچسب را نیز در نظر بگیرد و با اعمال منظم‌سازی تنک بر روی وزن‌ها، انتخاب ویژگی را به صورت مستقیم و درون‌مدلی انجام دهد. مقاله حاضر با هدف پر کردن این شکاف، چارچوبی جدید برای انتخاب ویژگی نیمه‌نظارتی تنک مبتنی بر خودرمزنگار گراف ارائه می‌دهد. جدول (۱) مقایسه روش پیشنهادی با روش‌های پیشین را نشان می‌دهد.

۳- روش پیشنهادی

در این بخش، چارچوب جدیدی برای انتخاب ویژگی نیمه‌نظارتی مبتنی بر ترکیب خودرمزگذار تک لایه و گراف طیفی نیمه‌نظارتی[1] (**GASFS**) ارائه می‌شود. این روش ساختار ذاتی داده یا به اصطلاح ساختار داده گسترده را با استفاده از یک خودرمزنگار و ساختار محلی داده به همراه اطلاعات برچسب را با استفاده از گراف طیفی نیمه‌نظارتی حفظ می‌کند. بنابراین، هدف اصلی این روش، حفظ ساختار داده در سه سطح مختلف است:

- **ساختار داده گسترده**[2]: این ساختار توسط خودرمزگذار تک لایه حفظ می‌شود. خودرمزنگار با یادگیری بازسازی داده‌ها، تلاش می‌کند تا اطلاعات مهم و ساختار کلی داده‌ها را در فضای نهان (لایه پنهان) خود فشرده کند.
- **ساختار داده محلی**[3]: این ساختار توسط گراف طیفی نیمه‌نظارتی حفظ می‌شود. گراف طیفی با مدل‌سازی روابط محلی بین نمونه‌های داده، اطلاعات مربوط به همسایگی و شباهت بین داده‌ها را در نظر می‌گیرد.
- **اطلاعات برچسب داده**: اطلاعات برچسب داده‌های موجود نیز در فرایند انتخاب ویژگی با استفاده از گراف طیفی نیمه‌نظارتی لحاظ می‌شود. این امر به روش پیشنهادی کمک می‌کند تا ویژگی‌هایی را انتخاب کند که علاوه بر حفظ ساختار داده، با اطلاعات برچسب نیز سازگار باشند.

برخلاف برخی رویکردهای موجود در انتخاب ویژگی نیمه‌نظارتی که از محدودیت متعامد[4] برروی ماتریس تبدیل استفاده می‌کنند، روش پیشنهادی از این محدودیت صرف‌نظر می‌کند. اعمال محدودیت متعامد می‌تواند منجر به محدود شدن فضای جستجو شده و درنتیجه، یافتن راه‌حل بهینه را دشوارتر می‌کند [۲۳]. با حذف این محدودیت، روش پیشنهادی می‌تواند فضای جستجوی بزرگ‌تری را بررسی کند و به احتمال زیاد به راه‌حل‌های بهتری دست یابد. به عبارت دیگر، روش پیشنهادی با عدم اعمال محدودیت‌های متعامد بر ماتریس تبدیل، انعطاف‌پذیری بیشتری در یادگیری ویژگی‌های بهینه دارد.

در بخش‌های بعدی مقاله، جزئیات بیشتری در مورد نحوه ساخت گراف طیفی، نحوه ترکیب آن با خودرمزنگار تک لایه و نحوه بهینه‌سازی مدل ارائه خواهیم داد. همچنین، نتایج آزمایش‌ها و مقایسه روش پیشنهادی با روش‌های موجود نیز ارائه خواهد شد.

۳-۱ تابع هدف پیشنهادی

تابع هدف پیشنهادی از سه بخش اصلی تشکیل شده است که هر کدام هدف خاصی را دنبال می‌کنند:

- **بازسازی**[5]: این بخش مبتنی‌بر خودرمزنگار تک لایه است و هدف آن حفظ ساختار داده گسترده است. خودرمزنگار با بهینه‌سازی یک تابع خطای بازسازی، تلاش می‌کند تا خطای بین داده‌های ورودی **(x)** و داده‌های بازسازی‌شده **(x')** را به حداقل برساند:

$$\mathcal{L}(\Theta) = \frac{1}{2n}\sum_{i=1}^{n}\|X^{(i)} - h(X^{(i)};\Theta)\|_2^2 = \frac{1}{2n}\| X - h(X;\Theta)\|_F^2 \quad (۳)$$

که

$$h(X^{(i)};\Theta) = \sigma(W_2 \cdot \sigma(W_1 X^{(i)} + b_1) + b_2)$$

و

$$\Theta = [W_1, W_2, b_1, b_2]$$

[1] Graph Autoencoder-based Semi-Supervised Feature Selection
[2] Broad Data Structure
[3] Local Data Structure
[4] Orthogonality Constraint
[5] Reconstruction Term

همان‌طور که اشاره شده است، از آنجا که ماتریس وزن W_1 مستقیماً روی داده ورودی اعمال می‌شود، هر ستون W_1 می‌تواند به عنوان معیاری برای اهمیت ویژگی مربوطه استفاده شود.

- **منظم‌سازی**[1]: این بخش برای ایجاد تنکی در ویژگی‌های انتخابی استفاده می‌شود. تنکی به این معناست که تعداد کمی از ویژگی‌ها دارای مقادیر غیر صفر باشند و بیشتر ویژگی‌ها مقدار صفر داشته باشند. این امر باعث می‌شود که ویژگی‌های مهم‌تر انتخاب شوند و از پیچیدگی مدل کاسته شود [۲۴]. در روش پیشنهادی از نرم $L_{2,1}$ برای این منظور استفاده شده است، که همان‌طور که پیش‌تر گفته شد همبستگی بین ویژگی‌ها را در انتخاب ویژگی لحاظ می‌کند. بنابراین عبارت تنظیم به صورت رابطه (٤) تعریف می‌شود:

$$R(\Theta) = \| W_1 \|_{2,1} \qquad (٤)$$

- **گراف طیفی نیمه‌نظارتی**: این بخش مبتنی‌بر نظریه گراف طیفی است و هدف آن حفظ ساختار محلی داده و استفاده از اطلاعات برچسب داده است [۲۵]. فرض اصلی این است که نقاط داده نزدیک در فضای ورودی، باید نمایش‌های مشابهی در فضای ویژگی داشته باشند [۲۶]. برای این منظور، یک گراف نیمه‌نظارتی در فضای داده ساخته می‌شود. در روش پیشنهادی از فاصله کسینوس به عنوان معیار تشابه استفاده شده است. ماتریس مجاورت گراف (**A**) به صورت رابطه (٥) تعریف می‌شود:

$$A(i,j) = \begin{cases} 1 & \text{if } Y^{(i)} = Y^{(j)} \\ \dfrac{X^{(i)T} X^{(j)}}{\| X^{(j)} \|_2 \| X^{(i)} \|_2} & \text{if } X^{(i)} \in N_k(X^{(j)}) \text{ or } X^{(j)} \in N_k(X^{(i)}) \text{ and } Y^{(i)} or Y^{(j)} \text{ are unknown} \\ 0 & \text{otherwise} \end{cases} \qquad (٥)$$

که $X^{(i)}$ و $X^{(j)}$ نمونه داده **i** ام و **j** ام، $Y^{(i)}$ و $Y^{(j)}$ برچسب نمونه داده **i** ام و **j** ام و $N_k(X^{(i)})$، مجموعه k–نزدیک‌ترین همسایه‌های نمونه $X^{(i)}$ اشاره دارد. ماتریس لاپلاسین L گراف G به صورت $L = D - A$ تعریف می‌شود، که D یک ماتریس قطری است و عناصر روی قطر اصلی آن به صورت

$$D(i,i) = \sum_{j=1}^{n} A(i,j)$$

تعریف می‌شود. برای حفظ ساختار محلی داده‌ها در زیرفضای آموزش‌دیده، عبارت (٦) کمینه می‌شود:

$$\begin{aligned} G(\Theta) &= \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \| Y^{(i)} - Y^{(j)} \|_2^2 \, A(i,j) G(\Theta) \\ &= \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} (Y^{(i)T}Y^{(i)} - Y^{(i)T}Y^{(j)} - Y^{(j)T}Y^{(i)} + Y^{(j)T}Y^{(j)}) A(i,j) G(\Theta) \\ &= \sum_{i=1}^{n} Y^{(i)T}Y^{(i)} D_{ii} - \sum_{i=1}^{n}\sum_{j=1}^{n} Y^{(i)T}Y^{(j)} A(i,j) \\ &= \mathrm{Tr}(Y(\Theta)DY(\Theta)^T) - \mathrm{Tr}(Y(\Theta)AY(\Theta)^T) = \mathrm{Tr}(Y(\Theta)LY(\Theta)^T) \end{aligned} \qquad (٦)$$

که $\mathrm{Tr}(.)$ عملگر اثر و $Y^{(i)}(\Theta) = \sigma(W_1 X^{(i)} + b_1)$ نمایش نهفته (ویژگی‌های فشرده‌شده) سس داده **i** بعد از عبور از لایه رمزگذار است. این عبارت تضمین می‌کند که اگر دو نقطه داده $X^{(i)}$ و $X^{(j)}$ در فضای داده اصلی نزدیک باشند، نمایش‌های کم‌بعد آن‌ها $Y^{(i)}$ و $Y^{(j)}$ نیز در فضای تعبیه‌شده نزدیک باشند. بنابراین، با ترکیب سه بخش تابع هدف انتخاب ویژگی مبتنی بر خودرمزنگار تک لایه، عبارت منظم‌سازی و گراف نیمه‌نظارتی حفظ ساختار داده‌های محلی، تابع هدف روش پیشنهادی را می‌توان به صورت کمینه‌سازی رابطه (٧) با توجه به پارامترهای $\Theta = [W_1, W_2, b_1, b_2]$ نوشت:

$$\begin{aligned} \hat{\Theta} &= \arg\min_{\Theta} F(\Theta) = \arg\min_{\Theta} \mathcal{L}(\Theta) + R(\Theta) + g(\Theta) \\ &= \arg\min_{\Theta} \left[\frac{1}{2n} \| X - h(X;\Theta) \|_F^2 + \lambda \| W_1 \|_{2,1} + \gamma Tr(Y(\Theta)LY(\Theta)^T) \right] \end{aligned} \qquad (٧)$$

که در آن λ و γ پارامترهای تنظیم هستند که وزن هر بخش را در تابع هدف تعیین می‌کنند. همچنین همان‌طور که در بخش (۲-۱) اشاره شد، W_1، b_1، W_2، b_2 به‌ترتیب به ماتریس وزن و بردار بایاس رمزگذار و رمزگشای خودرمزنگار اشاره دارند. انتخاب

[1] Regularization Term

ویژگی با استفاده از تابع امتیاز $\| \mathbf{W}_1^{(q)} \|_2$ براساس ماتریس وزن b_2 از $\hat{\Theta}$ انجام می‌شود.

۳-۲ بهینه‌سازی

این بخش از مقاله، نحوه محاسبه گرادیان‌های تابع هزینه را نسبت به پارامترهای مدل (وزن‌ها و بایاس‌ها) با استفاده از الگوریتم پس‌انتشار خطا توضیح می‌دهد. هدف از این محاسبات، به‌روزرسانی پارامترهای مدل در طی فرآیند آموزش با استفاده از روش‌های بهینه‌سازی مانند گرادیان کاهشی است.

گرادیان‌های مربوط به عبارت هزینه $\mathcal{L}(\Theta)$ با استفاده از الگوریتم پس‌انتشار به دست آمده و پارامترهای مدل در طی فرآیند آموزش به‌روزرسانی می‌شود [۱۸]. گرادیان‌های حاصل به صورت رابطه (۸) هستند:

$$\begin{aligned}
&\frac{\partial L(\Theta)}{\partial W_1} = \frac{1}{n}\Delta_2 X^T \\
&\frac{\partial L(\Theta)}{\partial W_2} = \frac{1}{n}\Delta_3 Y^T \\
&\frac{\partial L(\Theta)}{\partial b_1} = \frac{1}{n}\sum_{i=1}^{n}\Delta_2^{(i)} = \frac{1}{n}\Delta_2 1 \\
&\frac{\partial L(\Theta)}{\partial b_2} = \frac{1}{n}\sum_{i=1}^{n}\Delta_3^{(i)} = \frac{1}{n}\Delta_3 1
\end{aligned} \qquad (۸)$$

که $\boldsymbol{\Delta_2 = (W_2^T \Delta_3) \odot Y \odot (1 - Y)}$ خطای لایه مخفی، $\boldsymbol{\Delta_3 = (\hat{X} - X) \odot \hat{X} \odot (1 - \hat{X})}$ خطای لایه خروجی و $\odot$ عملگر ضرب عنصری است.

مشتق عبارت منظم‌سازی برای ستون $\mathbf{i}$ ام که ($\mathbf{W}_1^{(i)} = 0$ $i = 1, 2, \ldots, d$) وجود ندارد. در اینجا:

$$\frac{\partial R(\Theta)}{\partial W_1} = \lambda W_1 U \qquad (۹)$$

که $U \in \mathbb{R}^{d \times d}$ یک ماتریس قطری است که عناصر قطر اصلی به صورت رابطه (۱۰) تعریف می‌شود [۱۸]:

$$U(i,i) = \begin{cases} (\|W_1^{(i)}\|_2 + \epsilon)^{-1} & \|W_1^{(i)}\|_2 \neq 0 \\ 0 & otherwise \end{cases} \qquad (۱۰)$$

مقدار کوچک ϵ معمولاً برای جلوگیری از تقسیم بر صفر افزوده می‌شود [۱۹]. بنابراین مشتق عبارت منظم‌سازی نسبت به پارامترها به صورت رابطه (۱۱) تعریف می‌شود:

$$\begin{aligned}
&\frac{\partial R(\Theta)}{\partial W_1} = \lambda W_1 U \\
&\frac{\partial R(\Theta)}{\partial W_2} = 0 \\
&\frac{\partial R(\Theta)}{\partial b_1} = 0 \\
&\frac{\partial R(\Theta)}{\partial b_2} = 0
\end{aligned} \qquad (۱۱)$$

مشتق عبارت گراف نیمه‌نظارتی حفظ ساختار محلی داده‌ها نسبت به پارامترها به صورت رابطه (۱۲) تعریف می‌شود:

$$\begin{aligned}
&\frac{\partial G(\Theta)}{\partial W_1} = 2\gamma(YL \odot Y \odot (1-Y))X \\
&\frac{\partial G(\Theta)}{\partial b_1} = 2\gamma(YL \odot Y \odot (1-Y))\mathbf{1} \\
&\frac{\partial G(\Theta)}{\partial W_2} = 0 \\
&\frac{\partial G(\Theta)}{\partial b_2} = 0
\end{aligned} \qquad (۱۲)$$

درنهایت تمام مشتقات فوق برای هر پارامتر ترکیب شده تا گرادیان کل تابع هدف به دست آید:

$$\begin{aligned}
&\frac{\partial F(\Theta)}{\partial W_1} = \frac{1}{n}\Delta_2 X^T + \lambda W_1 U + 2\gamma(YL \odot Y \odot (1-Y))X^T \\
&\frac{\partial F(\Theta)}{\partial W_2} = \frac{1}{n}\Delta_3 Y \\
&\frac{\partial F(\Theta)}{\partial b_1} = \frac{1}{n}\Delta_2 \mathbf{1} + 2\gamma(YL \odot Y \odot (1-Y))\mathbf{1} \\
&\frac{\partial F(\Theta)}{\partial b_2} = \frac{1}{n}\Delta_3 \mathbf{1}
\end{aligned} \qquad (۱۳)$$

از الگوریتم (۱) برای حل روش پیشنهادی استفاده شده است.

الگوریتم ۱: الگوریتم GASFS

ورودی:

مجموعه داده با ابعاد بالا $\mathbf{X} = \left[\mathbf{X}^{(1)}, \mathbf{X}^{(2)}, \ldots, \mathbf{X}^{(n)}\right] \in R^{d \times n}$

اندازه همسایگی k

اندازه لایه پنهان m

پارامترهای تنظیم λ و γ

تعداد ویژگی‌هایی که باید نگه‌داشته شوند n_f

مرحله ۱: ساخت گراف

ساخت گراف k–نزدیک‌ترین همسایه G با ماتریس مجاورت A مطابق با رابطه (۱۵)

محاسبه ماتریس لاپلاسین L گراف G از ماتریس مجاورت A

مرحله ۲: بهینه‌سازی تابع هدف

بهینه‌سازی تابع هدف رابطه (۱۷) با استفاده از روش شرح داده شده در بخش ۳-۲

مرحله ۳: انتخاب ویژگی

محاسبه امتیاز $GAFS(p) = \left\| \mathbf{W}_1^{(p)} \right\|_2$ $(p=1,2,\ldots,d)$ برای هر ویژگی

ویژگی‌ها براساس امتیازها به صورت نزولی مرتب شده و اندیس n_f ویژگی اول برگردانده می‌شود.

٤- ارزیابی

برای ارزیابی جامع روش پیشنهادی، از شش مجموعه داده استاندارد از مخزن معتبر UCIاستفاده شده است. این مجموعه داده‌ها به‌دلیل تنوع در تعداد نمونه‌ها، ویژگی‌ها و کلاس‌ها، برای ارزیابی عملکرد روش‌های انتخاب ویژگی مناسب هستند. جزئیات این مجموعه داده‌ها در جدول (۲) آمده است [۲۲, ۲۳, ۲٤].

روش پیشنهادی برای انتخاب ویژگی نیمه‌نظارتی با استفاده از داده‌های برچسب‌دار و بدون برچسب، برروی مجموعه داده‌های گفته شده ارزیابی و با روش‌های رایج (SFS[۱۵]، GS²FS-L21 [۱٤]، FSNM [۱۲]، SFSN [۱۳] و S²DFS [۳۰]) مقایسه شده است.

به‌صورت تصادفی ۷۰ درصد داده‌ها برای آزمون و ۳۰ درصد برای تست در نظر گرفته‌شده است، همچنین داده‌ها به سه دسته آموزشی بدون برچسب، آموزشی برچسب‌دار و آزمون تقسیم شده و آزمایش‌ها به صورت جداگانه بر روی هر مجموعه داده سه بار تکرار شده است. ۵۰ درصد از داده‌ها به صورت تصادفی انتخاب و به‌عنوان داده بدون برچسب درنظر گرفته شده است.

جدول (۲): مجموعه داده‌های استفاده شده

مجموعه داده	تعداد ویژگی	تعداد نمونه	تعداد کلاس
ORL	۱۰۲٤	٤۰۰	٤۰
ATT	۱۰۲٤	٤۰۰	٤۰
WBCD	۱۰	۳۹۹	۲
WDBC	۳۰	٥٦۹	۲
QSAR	٥۱	۱۰٥٥	۲
پارکینسون	۱۹٥	۲۲	۲

برای ارزیابی کیفیت ویژگی‌های انتخاب شده، از دسته‌بندی ماشین بردار پشتیبان (SVM) و k–نزدیک‌ترین همسایه (KNN) استفاده شده و معیار دقت به‌عنوان معیار ارزیابی در نظر گرفته شده است. پارامترهای مربوطه با استفاده از روش اعتبارسنجی متقاطع دسته‌ای–۱۰ (10-Fold Cross Validation) تنظیم شده و نتایج به صورت میانگین و انحراف معیار گزارش شده‌اند. مقادیر بهینه پارامترهای منظم‌سازی $\boldsymbol{\lambda}$ و گراف طیفی نیمه‌نظارتی γ در روش پیشنهادی رابطه (۷) با استفاده از روش اعتبارسنجی متقاطع دسته‌ای–۱۰ از مجموعه {۰.۰۱، ۰.۱، ۱، ۱۰} تعیین شده است.

جدول (۳) و (٤) نتایج دقت طبقه‌بندی روش‌های مرسوم انتخاب ویژگی و روش پیشنهادی با استفاده از ماشین بردار پشتیبان و k–نزدیک‌ترین همسایه را نشان می‌دهد. نتایج آزمایش برروی شش مجموعه داده حاکی از برتری روش پیشنهادی در مقایسه با سایر روش‌های انتخاب ویژگی در هر دو الگوریتم طبقه‌بندی است.

همچنین برای بررسی اعتبار نتایج، از آزمون Wilcoxon signed-rank جهت مقایسه روش پیشنهادی با پنج روش مرجع در دو طبقه‌بند SVM و KNN استفاده شد. همان‌طور که در جدول‌های (٤) و (٥) نشان داده شده است، در تمامی مقایسه‌ها مقدار آماره آزمون برابر با $W = ۰$ است؛ این نتیجه نشان می‌دهد که در هیچ‌یک از مجموعه‌داده‌ها روش‌های مرجع برتر از روش پیشنهادی نبوده‌اند. همچنین مقادیر p به‌دست‌آمده برابر با ۰/۰۳۱۲ و کمتر از ۰.۰٥ است، بنابراین اختلاف عملکرد روش پیشنهادی نسبت به رقبا از نظر آماری معنادار است. علاوه بر این، مقدار اندازه اثر محاسبه‌شده برابر با $r = ۰/۸۷۹$ بوده که نشان‌دهنده یک اثر بسیار بزرگ است؛ به عبارت دیگر، برتری روش پیشنهادی نه تنها از نظر

آماری معتبر است، بلکه شدت این برتری نیز چشمگیر و قابل‌توجه است.

جدول (۳): دقت طبقه‌بندی (میانگین ±انحراف معیار) ماشین بردار پشتیبان [۲۲, ۲۳, ۲۴]

مجموعه داده	S^2DFS	SFSN	FSNM	$GS^2FS\text{-}L_{21}$	SFS	GASFS
ORL	۰/۰۳۳ ±۰/۵۶۱۰	۰/۰۴۷± ۰/۴۰۵۰	۰/۰۲۵ ±۰/۶۷۲۰	۰/۰۹۴±۰/۶۳۰۰	۰/۰۱۶ ±۰/۶۶۶۰	۰/۷۸۰۰±۰/۰۳۴
ATT	۰/۰۵۰ ± ۰/۵۸۳۰	۰/۰۵۲± ۰/۴۳۰۰	۰/۱۰۶ ±۰/۶۳۶۰	۰/۱۱۵ ± ۰/۶۱۶۰	۰/۰۵۰ ± ۰/۶۸۳۰	۰/۸۸۳۰±۰/۰۲۳
WBCD	۰/۹۵۷۰±۰/۰۰۶	۰/۹۵۹۰±۰/۰۰۸	۰/۹۴۰۰±۰/۰۱۴	۰/۹۶۷۰±۰/۰۰۶	۰/۹۶۰۰±۰/۰۱۵	۰/۹۸۲۰±۰/۰۰۸
WDBC	۰/۹۱۹۰±۰/۰۱۱	۰/۹۴۰۰±۰/۰۲۴	۰/۹۴۰۰±۰/۰۰۱۱	۰/۹۴۰۰±۰/۰۱۴	۰/۹۱۰۰±۰/۰۲۷	۰/۹۷۰۰±۰/۰۱۶
QSAR	۰/۸۰۰۰±۰/۰۱۳	۰/۷۶۳۰±۰/۰۱۷	۰/۷۸۹۰±۰/۰۱۹	۰/۸۰۹۰±۰/۰۱۱	۰/۷۸۰۰±۰/۰۰۵	۰/۸۱۰۰±۰/۰۴۱
پارکینسون	۰/۸۴۸۰±۰/۰۳۴	۰/۸۰۰۰±۰/۰۲۶	۰/۸۲۰۰±۰/۰۳۹	۰/۸۶۰۰±۰/۰۱۷	۰/۸۱۰۰±۰/۰۱۷	۰/۹۳۰۰±۰/۰۰۰

شایان ذکر است که مقادیر آماره Wilcoxon، p-value و اندازه اثر در هر دو طبقه‌بندSVM و KNN یکسان به دست آمدند. این موضوع بیانگر پایداری و استقلال عملکرد روش پیشنهادی از نوع طبقه‌بند است.

جدول (٤): دقت طبقه‌بندی (میانگین ±انحراف معیار) k-نزدیک‌ترین همسایه [۲۲, ۲۳, ۲۴]

مجموعه داده	S^2DFS	SFSN	FSNM	$GS^2FS\text{-}L_{21}$	SFS	GASFS
ORL	۰/۶۰۲۸±۰/۰۶۰	۰/۴۷۲۲±۰/۰۱۲	۰/۶۰۲۸±۰/۱۵۰	۰/۶۷۲۲± ۰/۰۳	۰/۷۱۱۰±۰/۰۵	۰/۷۵۶۱±۰/۰۳۴
ATT	۰/۶۷۳۰ ±۰/۰۶۱	۰/۵۳۸۹±/۰۶۱	۰/۷۳۶۱±۰/۰۳۹	۰/۷۵۸۳ ±۰/۰۲	۰/۷۵۵۶± ۰/۰۳۳	۰/۹۲۵۰±۰/۰۲۳
WBCD	۰/۰۲۴±۰/۹۰۵۲	۰/۹۶۰۰±۰/۰۱۳	۰/۹۵۱۰±۰/۰۱۰	۰/۹۴۶۰±۰/۰۱۲	۰/۹۶۱۰±۰/۰۱۴	۰/۹۷۷۴±۰/۰۰۸
WDBC	۰/۹۰۵۰±۰/۰۳۲	۰/۹۳۸۰±۰/۰۱۱	۰/۹۳۴۰±۰/۰۱۱	۰/۹۲۹۰±۰/۰۱۶	۰/۹۲۱۰±۰/۰۱۶	۰/۹۴۹۸±۰/۰۱۶
QSAR	۰/۷۸۰۰±۰/۰۰۵	۰/۷۹۳۰±۰/۰۲۱	۰/۷۸۹۰±۰/۰۱۹	۰/۷۶۳۰±۰/۰۱۷	۰/۷۹۰۰±۰/۰۱۳	۰/۸۲۴۸±۰/۰۴۱
پارکینسون	۰/۷۹۷۰/±۰/۰۲۹	۰/۸۰۸۰±۰/۰۴۳	۰/۸۰۸۰±۰/۰۲۶	۰/۸۰۲۰±۰/۰۲۰	۰/۸۰۸±۰۰/۰۴۳	۰/۸۹۶۶±۰/۰۰۰

جدول (۵): نتایج آزمون Wilcoxon بین روش پیشنهادی و روش‌های مرجع براساس دو طبقه‌بند SVM و KNN

روش مقایسه‌ای	W	p-value	اندازه اثر (r)
SFS	۰/۰	۰/۰۳۲۱	۰/۸۷۹
$GS^2FS\text{-}L_{21}$	۰/۰	۰/۰۳۲۱	۰/۸۷۹
FSNM	۰/۰	۰/۰۳۲۱	۰/۸۷۹
SFSN	۰/۰	۰/۰۳۲۱	۰/۸۷۹
S^2DFS	۰/۰	۰/۰۳۲۱	۰/۸۷۹

این نتایج نشان می‌دهد که استفاده از اطلاعات داده گسترده به همراه ساختار محلی داده‌های برچسب‌دار و بدون برچسب منجر به بهبود فرآیند انتخاب ویژگی می‌شود و ویژگی‌های مناسب‌تری انتخاب می‌شوند.

٤-۱ مطالعه همگرایی

در این بخش عملکرد همگرایی الگوریتم ۱ به‌صورت تجربی ارزیابی می‌شود. آزمایش‌ها برروی شش مجموعه داده مختلف انجام شده است. شکل (۱)، روند همگرایی مقادیر تابع هدف تعریف شده در رابطه (۷) را، که با استفاده از الگوریتم (۱) و برمبنای خودرمزنگار محاسبه شده‌اند، به نمایش می‌گذارد. همان‌گونه که مشاهده می‌شود، مقدار تابع هدف در تکرارهای اولیه با شیب نسبتاً زیادی کاهش یافته است که بیانگر سرعت بالای همگرایی در مراحل آغازین آموزش است. پس از چندین تکرار،

تغییرات تابع هدف تدریجاً کاهش یافته و الگوریتم به یک مقدار پایدار نزدیک می‌شود. این رفتار نشان‌دهنده همگرایی یکنواخت و بدون نوسان شدید است که مطلوبیت الگوریتم را از نظر پایداری آموزش تأیید می‌کند و حاکی از کارآمدی و اثربخشی الگوریتم (۱) در بهینه‌سازی تابع هدف مورد نظر می‌باشد.

۴-۲ پیچیدگی زمانی الگوریتم پیشنهادی

پیچیدگی زمانی روش پیشنهادی را می‌توان بر اساس مراحل اصلی الگوریتم تحلیل نمود:

ساخت گراف k–نزدیک‌ترین همسایه: برای یک مجموعه‌داده با n نمونه و dویژگی، محاسبه شباهت بین هر جفت نمونه هزینه‌ای در حد $O(n^2 d)$دارد. با استفاده از روش‌های تقریب، این هزینه می‌تواند به $O(n \log n \cdot d)$کاهش یابد.

آموزش خودرمزنگار تک‌لایه: در هر تکرار آموزش، هزینه محاسبه بازسازی و گرادیان برابر با O(ndm) است که در آن mتعداد نورون‌های لایه پنهان است. اگر تعداد تکرارهای الگوریتم بهینه‌سازی را T در نظر بگیریم، هزینه کل این بخش برابر با O(T·ndm)خواهد بود.

اعمال منظم‌سازی تنک: محاسبه گرادیان مربوط به عبارت منظم‌سازی هزینه‌ای در حدود O(dm) در هر تکرار دارد که در مقایسه با هزینه آموزش خودرمزنگار قابل صرف‌نظر کردن است.

عبارت گراف طیفی نیمه‌نظارتی: محاسبه عبارت لاپلاسین و به‌روزرسانی گرادیان‌ها در هر تکرار هزینه‌ای در حدود $O(nm^2)$ دارد. این هزینه در عمل معمولاً از هزینه آموزش خودرمزنگار کوچک‌تر است زیرا d≫m.

بنابراین، پیچیدگی زمانی کلی الگوریتم پیشنهادی را می‌توان به صورت رابطه (۱۴) خلاصه کرد:

$$O(n^2 d + T \cdot ndm + T \cdot nm^2) \qquad (۱۴)$$

در داده‌های با ابعاد بالا، جمله غالب O(T·ndm)است که مربوط به آموزش خودرمزنگار می‌باشد. با این حال، با انتخاب اندازه نهان d≫m پیچیدگی به طور قابل‌توجهی کاهش می‌یابد. از این رو، روش پیشنهادی از نظر محاسباتی کارا بوده و امکان به‌کارگیری آن در مجموعه‌داده‌های متوسط و بزرگ فراهم است. همچنین می‌توان با استفاده از پیاده‌سازی‌های موازی (GPU) و روش‌های تقریبی در ساخت گراف، زمان اجرای الگوریتم را بیشتر کاهش داد.

۵- نتیجه‌گیری

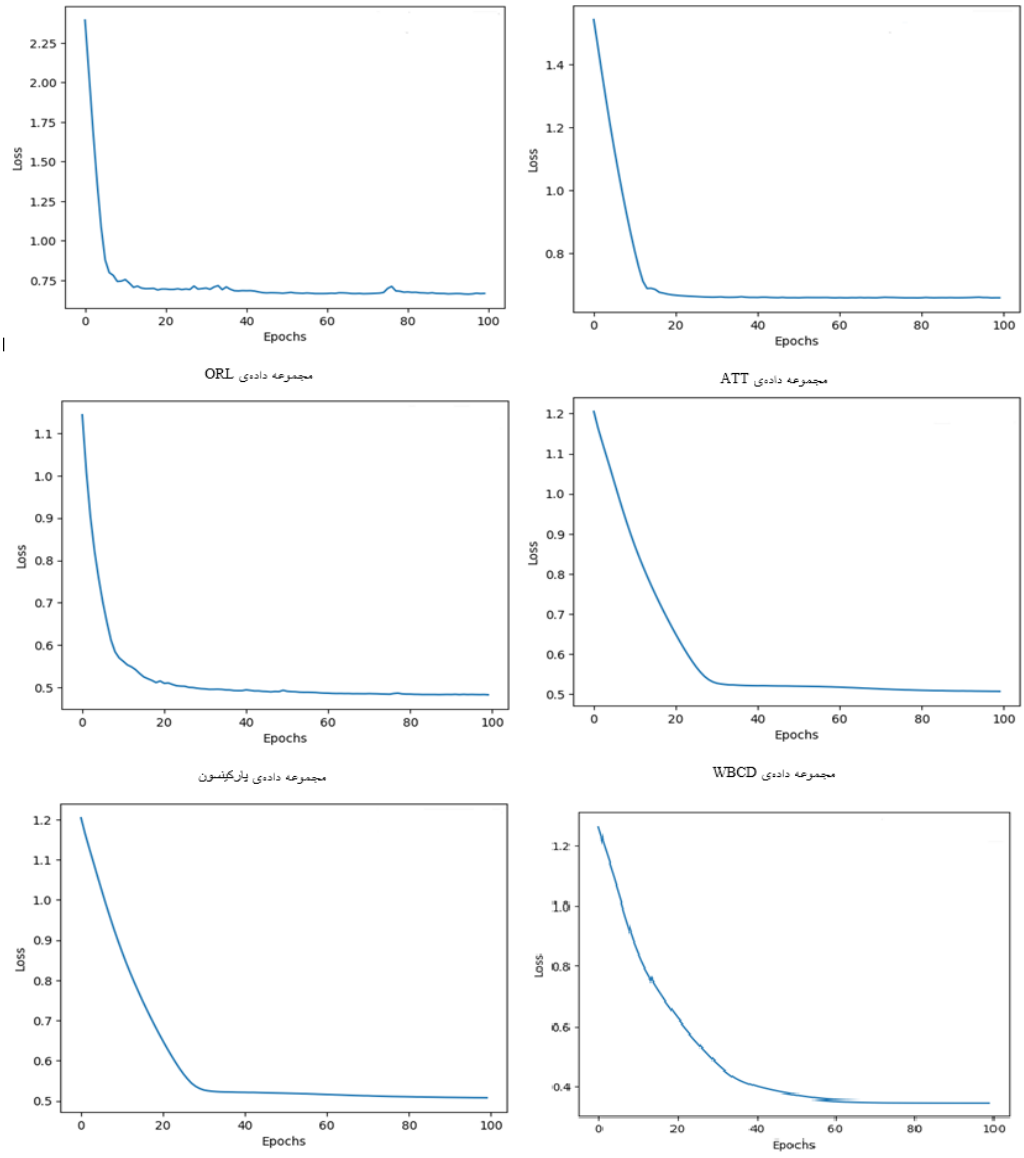
در این مقاله، یک چارچوب جدید انتخاب ویژگی نیمه‌نظارتی مبتنی‌بر ترکیب خودرمزگذار و گراف طیفی ارائه گردید. هدف اصلی این روش، بهره‌گیری هم‌زمان از اطلاعات موجود در داده‌های برچسب‌دار و بدون‌برچسب برای انتخاب زیرمجموعه‌ای بهینه از ویژگی‌ها بود. روش پیشنهادی با استفاده از خودرمزگذار تک‌لایه، ساختار کلی و گسترده داده‌ها را حفظ می‌کند و با بهره‌گیری از گراف طیفی نیمه‌نظارتی، ساختار محلی داده‌ها و اطلاعات برچسب‌های موجود را در فرآیند انتخاب ویژگی لحاظ می‌نماید.

یکی از ویژگی‌های متمایزکننده روش پیشنهادی، استفاده از نرم- $L_{2,1}$ برای ایجاد تنکی در ماتریس وزن لایه اول خودرمزگذار است که منجربه انتخاب ویژگی‌های متمایز کننده می‌شود. همچنین، برخلاف برخی روش‌های پیشین، روش پیشنهادی از اعمال محدودیت متعامد بر روی ماتریس تبدیل صرف‌نظر می‌کند که این امر انعطاف‌پذیری بیشتری به مدل داده و فضای جستجوی بزرگ‌تری را برای یافتن ویژگی‌های بهینه فراهم می‌آورد.

ارزیابی‌های تجربی بر روی شش مجموعه‌داده استاندارد از مخزن UCI نشان داد که روش پیشنهادی با استفاده از دو طبقه‌بند ماشین بردار پشتیبان و k–نزدیک‌ترین همسایه در شش مجموعه‌داده دقت بالاتری نسبت به بهترین روش‌های مقایسه‌ای کسب کرده است. علاوه بر این، آزمونWilcoxon signed-rank با سطح اطمینان ۹۵٪ برتری معنادار روش پیشنهادی نسبت به تمامی روش‌های مرجع را تأیید نمود؛ به‌طوری‌که مقدار آماره آزمون برابر W = ۰و اندازه اثر محاسبه‌شده r =۰/۸۷۹ نشان‌دهنده یک اثر بسیار بزرگ بود. این نتایج بیانگر آن است که اهداف مطرح‌شده در مقدمه، شامل حفظ ساختار سراسری و محلی داده‌ها، بهره‌گیری از اطلاعات برچسب و دستیابی به انتخاب ویژگی تنک و کارآمد، به‌طور کامل محقق شده‌اند. علاوه بر این، تحلیل همگرایی الگوریتم بهینه‌سازی شکل (۱) نشان داد که مقدار تابع هدف در مراحل اولیه با سرعت زیادی کاهش یافته و سپس به‌صورت یکنواخت به مقدار پایدار

همگرا می‌شود. این روند بیانگر همگرایی سریع و پایداری الگوریتم و درنتیجه تأیید کارایی محاسباتی روش پیشنهادی است. با وجود این دستاوردها، روش پیشنهادی دارای محدودیت‌هایی نیز هست. نخست، ساخت گراف k–نزدیک‌ترین همسایه در داده‌های بسیار بزرگ از نظر زمانی پرهزینه است، هرچند می‌توان از نسخه‌های تقریبی برای کاهش این هزینه استفاده کرد. یکی دیگر از محدودیت‌های روش پیشنهادی وابستگی نسبی آن به پارامترهای تنظیم است. در این پژوهش، مقادیر این پارامترها از میان چند مقدار کاندیدا انتخاب شده است.

به‌عنوان مسیر آینده، می‌توان بر توسعه نسخه‌های مقیاس‌پذیر و بهینه‌سازی‌شده برای داده‌های حجیم یا جریانی و همچنین به‌کارگیری چارچوب‌های عمیق‌تر نظیر CNN و Transformer برای داده‌های پیچیده‌تر تمرکز کرد.



شکل (۱): منحنی‌های همگرایی الگوریتم ۱ با توجه به تابع هدف

References

[1] G. Chandrashekar and F. Sahin, "A survey on feature selection methods," Computers & Electrical Engineering, vol. 40, no. 1, pp. 16–28, Jan. 2014, doi: 10.1016/j.compeleceng.2013.11.024.

[2] Y. Hu, Y. Zhang, and D. Gong, "Multiobjective Particle Swarm Optimization for Feature Selection With Fuzzy Cost," IEEE Trans Cybern, vol. 51, no. 2, pp. 874–888, Feb. 2021, doi: 10.1109/TCYB.2020.3015756.

[3] W. Zhong, X. Chen, F. Nie, and J. Z. Huang, "Adaptive discriminant analysis for semi-supervised feature selection," Inf Sci (N Y), vol. 566, pp. 178–194, Aug. 2021, doi: 10.1016/j.ins.2021.02.035.

[4] G. Roffo, S. Melzi, U. Castellani, A. Vinciarelli, and M. Cristani, "Infinite Feature Selection: A Graph-based Feature Filtering Approach," IEEE Trans Pattern Anal Mach Intell, vol. 43, no. 12, pp. 4396–4410, Dec. 2021, doi: 10.1109/TPAMI.2020.3002843.

[5] R. Sheikhpour, M. A. Sarram, S. Gharaghani, and M. A. Z. Chahooki, "A Survey on semi-supervised feature selection methods," Pattern Recognit, vol. 64, pp. 141–158, Apr. 2017, doi: 10.1016/j.patcog.2016.11.003.

[6] T. Bhadra and S. Bandyopadhyay, "Supervised feature selection using integration of densest subgraph finding with floating forward–backward search," Inf Sci (N Y), vol. 566, pp. 1–18, Aug. 2021, doi: 10.1016/j.ins.2021.02.034.

[7] B. C. Love, "Comparing supervised and unsupervised category learning," Psychon Bull Rev, vol. 9, no. 4, pp. 829–835, Dec. 2002, doi: 10.3758/BF03196342.

[8] R. Zhang, H. Zhang, X. Li, and S. Yang, "Unsupervised Feature Selection With Extended OLSDA via Embedding Nonnegative Manifold Structure," IEEE Trans Neural Netw Learn Syst, vol. 33, no. 5, pp. 2274–2280, May 2022, doi: 10.1109/TNNLS.2020.3045053.

[9] J. E. van Engelen and H. H. Hoos, "A survey on semi-supervised learning," Mach Learn, vol. 109, no. 2, 2020, doi: 10.1007/s10994-019-05855-6.

[10] C. Shi, Q. Ruan, and G. An, "Sparse feature selection based on graph Laplacian for web image annotation," Image Vis Comput, vol. 32, no. 3, pp. 189–201, Mar. 2014, doi: 10.1016/j.imavis.2013.12.013.

[11] H. Barkia, H. Elghazel, and A. Aussem, "Semi-supervised Feature Importance Evaluation with Ensemble Learning," in 2011 IEEE 11th International Conference on Data Mining, IEEE, Dec. 2011, pp. 31–40. doi: 10.1109/ICDM.2011.129.

[12] F. Nie, H. Huang, X. Cai, and C. Ding, "Efficient and robust feature selection via joint ℓ2, 1-norms minimization," Adv Neural Inf Process Syst, vol. 23, 2010.

[13] L. Wang and S. Chen, "$ l_ {2, p} $ Matrix Norm and Its Application in Feature Selection," arXiv preprint arXiv:1303.3987, 2013.

[14] R. Sheikhpour, M. A. Sarram, S. Gharaghani, and M. A. Z. Chahooki, "A robust graph-based semi-supervised sparse feature selection method," Inf Sci (N Y), vol. 531, pp. 13–30, Aug. 2020, doi: 10.1016/j.ins.2020.03.094.

[15] X. Li, Y. Zhang, and R. Zhang, "Semisupervised Feature Selection via Generalized Uncorrelated Constraint and Manifold Embedding," IEEE Trans Neural Netw Learn Syst, vol. 33, no. 9, pp. 5070–5079, Sep. 2022, doi: 10.1109/TNNLS.2021.3069038.

[16] G. Sampson, D. E. Rumelhart, J. L. McClelland, and The PDP Research Group, "Parallel Distributed Processing: Explorations in the Microstructures of Cognition," Language (Baltim), vol. 63, no. 4, p. 871, Dec. 1987, doi: 10.2307/415721.

[17] Y. Bengio, "Learning Deep Architectures for AI," Foundations and Trends® in Machine Learning, vol. 2, no. 1, pp. 1–127, 2009, doi: 10.1561/2200000006.

[18] S. Feng and M. F. Duarte, "Graph autoencoder-based unsupervised feature selection with broad and local data structure preservation," Neurocomputing, vol. 312, pp. 310–323, Oct. 2018, doi: 10.1016/j.neucom.2018.05.117.

[19] R. Shang, Z. Zhang, L. Jiao, C. Liu, and Y. Li, "Self-representation based dual-graph regularized feature selection clustering," Neurocomputing, vol. 171, pp. 1242–1253, Jan. 2016, doi: 10.1016/j.neucom.2015.07.068.

[20] P. Vincent, H. Larochelle, Y. Bengio, and P. A. Manzagol, "Extracting and composing robust features with denoising autoencoders," in Proceedings of the 25th International Conference on Machine Learning, 2008. doi: 10.1145/1390156.1390294.

[21] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," in 2nd International Conference on Learning Representations, ICLR 2014 - Conference Track Proceedings, 2014. doi: 10.61603/ceas.v2i1.33.

[22] A. Makhzani, J. Shlens, N. Jaitly, I. Goodfellow, and B. Frey, "Adversarial autoencoders," arXiv preprint arXiv:1511.05644, 2015.

[23] Z. Xu, X. Chang, F. Xu, and H. Zhang, "$L_{1/2}$ Regularization: A Thresholding Representation Theory and a Fast Solver," IEEE Trans Neural Netw Learn Syst, vol. 23, no. 7, pp. 1013–1027, Jul. 2012, doi: 10.1109/TNNLS.2012.2197412.

[24] X. Zhu, S. Zhang, Z. Jin, Z. Zhang, and Z. Xu, "Missing Value Estimation for Mixed-Attribute Data Sets," IEEE Trans Knowl Data Eng, vol. 23, no. 1, pp. 110–121, Jan. 2011, doi: 10.1109/TKDE.2010.99.

[25] D. Cai, C. Zhang, and X. He, "Unsupervised feature selection for multi-cluster data," in Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining, New York, NY, USA: ACM, Jul. 2010, pp. 333–342. doi: 10.1145/1835804.1835848.

[26] L. Bottou and V. Vapnik, "Local Learning Algorithms," Neural Comput, vol. 4, no. 6, pp. 888–900, Nov. 1992, doi: 10.1162/neco.1992.4.6.888.

[27] R. Sheikhpour, "Semi-supervised sparse feature selection based on Hessian regularization and Fisher discriminant analysis," Tabriz J. Electr. Eng., vol. 52, no. 2, pp. 125–135, 2022. doi: 10.22034/tjee.2022.15428. [In Persian]

[28] R. Sheikhpour, K. Berahmand, and S. Forouzandeh, "Hessian-based semi-supervised feature selection using generalized uncorrelated constraint," Knowl Based Syst, vol. 269, p. 110521, Jun. 2023, doi: 10.1016/j.knosys.2023.110521.

[29] R. Sheikhpour, "A local spline regression-based framework for semi-supervised sparse feature selection," Knowl Based Syst, vol. 262, p. 110265, Feb. 2023, doi: 10.1016/j.knosys.2023.110265.

[30] Z. Wang, F. Nie, L. Tian, R. Wang, and X. Li, "Discriminative Feature Selection via A Structured Sparse Subspace Learning Module," in Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, California: International Joint Conferences on Artificial Intelligence Organization, Jul. 2020, pp. 3009–3015. doi: 10.24963/ijcai.2020/416.

Journal of Applied And Basic Machine Intelligence Research (ABMIR)

Applied and Basic Machine Intelligence Research

Vol. 3, No. 1, Spring & Summer 2025, pp. 163 - 177

Semi-supervised Sparse Feature Selection based on Graph Autoencoder by Preservation of Broad and Local Data Structures

MohammadJavd Reezaei [1], MahdiAgha Sarram [2], Razieh Sheikhpour[3*]

[1]Phd candidate, Computer Engineering Department, Yazd University, Yazd, Iran
[2]Associate Professor, Computer Engineering Department, Yazd University, Yazd, Iran
[3] Associate Professor, Department of Computer Engineering, Faculty of Engineering, Ardakan University, Ardakan, Iran

Article Information

Original Research Paper

Received:
2025 July 9

Accepted:
2025 October 12

Keywords:
Semi-supervised, Feature selection, Auto encoder, Sparse models, $L_{2,1}$-norm

Corresponding Author*:
rsheikhpour@ardakan.ac.ir

Abstract

Processing and analyzing high-dimensional data is a significant challenge in many domains, and feature selection, as an effective dimension reduction method, plays a key role in improving the performance of machine learning models. Given that in the real world, labeling large volumes of data is costly and time-consuming, semi-supervised feature selection methods that can leverage valuable information from unlabeled data alongside labeled data have gained considerable importance. In this paper, a novel sparse semi-supervised feature selection framework is introduced, which simultaneously preserves the broad and local structures of data as well as the information from available labels. The proposed framework by optimizing a comprehensive objective function comprising an autoencoder reconstruction term, an L_(2,1)-norm regularization term for sparsity, and a term based on the semi-supervised spectral graph, selects an optimal subset of features. To solve this optimization problem, a gradient-based backpropagation algorithm is employed, and its convergence has been empirically investigated and confirmed. Extensive evaluations on six standard datasets and comparison of the results with several prominent previous methods demonstrate the significant superiority of the proposed framework in improving classification accuracy and selecting more effective features under semi-supervised conditions.

: 10.22034/ABMIR.2025.23363.1140

E-ISSN: 2821-2037