Journal of Applied And Basic Machine Intelligence Research (ABMIR)

پژوهش‌های نظری و کاربردی هوش ماشینی

سال چهارم، شماره دوم، پاییز و زمستان ۱۴۰۵، صفحات۲۰۴ - ۱۷۳

یک روش جدید تجزیه ماتریسی نامنفی استوار توزیعی مبتنی بر منظم‌ساز گراف دوتصادفی برای خوشه‌بندی داده

سید وفا بارخدا [۱]، فردین اخلاقیان طاب [۲*]

[۱] دانشجوی دکتری هوش مصنوعی و رباتیکز، دانشکده مهندسی، دانشگاه کردستان، سنندج، ایران

[۲] دانشیار گروه کامپیوتر، دانشکده مهندسی، دانشگاه کردستان، سنندج، ایران

مقاله پژوهشی

تاریخ دریافت:
۱۴۰۵/۰۴/۰۵

تاریخ پذیرش:
۱۴۰۵/۰۷/۰۸

کلیدواژه‌ها:
خوشه‌بندی داده، یادگیری استوار توزیعی، تجزیه ماتریسی استوار، تجزیه ماتریسی نامنفی، منظم‌ساز گراف دوتصادفی

نویسنده‌مسئول:
f.akhlaghian@uok.ac.ir

چکیده

تجزیه ماتریسی نامنفی (NMF) یکی از روش‌های پرکاربرد برای یادگیری بازنمایی داده و خوشه‌بندی داده‌های نامنفی است. با این حال، اکثر روش‌های موجود مبتنی بر فرض یک توزیع نوفه مشخص هستند و در مواجهه با داده‌های واقعی که معمولاً دارای نوفه‌های ناشناخته، ناهمگن یا ترکیبی هستند، با افت عملکرد مواجه می‌شوند. به‌منظور رفع این محدودیت، در این مقاله چارچوبی جدید با عنوان تجزیه ماتریسی نامنفی استوار توزیعی با منظم‌ساز گراف دوتصادفی (DRNMF-bS) ارائه می‌شود. هسته اصلی روش پیشنهادی بر پایه یک مدل تجزیه ماتریسی نامنفی استوار توزیعی نمونه‌محور است که در آن تابع هدف به‌صورت ترکیبی تطیقی از چندین تابع زیان متناظر با توزیع‌های مختلف نوفه تعریف شده و مسئله در قالب یک بهینه‌سازی حداقل–حداکثر بر روی مجموعه ابهام احتمالاتی فرمول‌بندی می‌شود. این چارچوب امکان یادگیری خودکار وزن توابع زیان و در نتیجه سازگاری پویا با شرایط نوفه‌ای مختلف را فراهم کرده و موجب افزایش پایداری مدل در برابر طیف گسترده‌ای از توزیع‌های نوفه و ترکیب آن‌ها می‌شود. علاوه بر این، به‌منظور بهره‌برداری مؤثر از ساختار هندسی داده‌ها، یک منظم‌ساز گراف دوتصادفی به مدل افزوده شده است که از طریق یادگیری ماتریس شباهت، اطلاعات محلی داده را به‌صورت درون‌زا در فرآیند تجزیه ادغام می‌کند. برای حل مسئله بهینه‌سازی حاصل، یک الگوریتم تکراری با تضمین کاهش تابع هدف ارائه شده است. نتایج آزمایش‌ها بر روی ۱۲ مجموعه‌داده و در مقایسه با ۱۰ روش مبتنی بر NMF نشان می‌دهد که روش پیشنهادی از نظر دقت خوشه‌بندی، مقاومت در برابر نوفه و قابلیت تعمیم، عملکردی برتر و پایدارتر نسبت به روش‌های رقیب ارائه می‌دهد.

: 10.22034/ABMIR.2026.24826.1262

E-ISSN: 2821-2037
/The Author 2026. 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

خوشه‌بندی یکی از وظایف اساسی در یادگیری بدون نظارت و یکی از تکنیک‌های کلیدی در یادگیری ماشین [۱] و داده‌کاوی است [۲]. هدف آن گروه‌بندی نمونه‌های داده‌ای به خوشه‌های معنادار است به‌گونه‌ای که نمونه‌های درون یک گروه دارای شباهت بالایی باشند و نمونه‌های بین گروه‌های مختلف از یکدیگر متمایز باشند [۳]. خوشه‌بندی در طیف گسترده‌ای از کاربردها ازجمله بینایی ماشین [٤]، تحلیل متون [٥]، بیوانفورماتیک [٦] و پردازش داده‌های چندحسگره [۷] مؤثر است. بااین‌حال، داده‌های واقعی اغلب دارای ابعاد بالا و ترکیبی از نماهای ناهمگن متعدد هستند که کشف ساختارهای پنهان معنادار را به‌ویژه در حضور نوفه، داده‌های پرت و توزیع‌های پیچیده داده، دشوار می‌کند [۸]. با افزایش مداوم حجم و تنوع داده‌ها، نیاز به روش‌های خوشه‌بندی که هم مؤثر و هم استوار در برابر نوفه باشند، بیشتر شده است. برای پاسخ به این چالش‌ها، رویکردهای متنوعی پیشنهاد شده‌اند که ازجمله آن‌ها می‌توان به الگوریتم‌های کلاسیک تکراری مانند K–میانگین [۹]، روش‌های مبتنی بر گراف مانند خوشه‌بندی طیفی [۱۰] و روش‌های تجزیه ماتریسی مانند تجزیه ماتریسی نامنفی [۱۱] اشاره کرد.

تجزیه ماتریسی نامنفی [۱۲][۱۳] به‌عنوان یک ابزار قدرتمند در بازیابی اطلاعات [۱٤] و بازشناسی الگو [۱٥] مطرح‌شده است، که عمدتاً به دلیل قابلیت تفسیر شهودی و معنادار آن شناخته می‌شود. NMF با تجزیه یک ماتریس داده نامنفی به دو عامل نامنفی کم‌رتبه، نمایشی جمع‌شونده و مبتنی بر اجزا ارائه می‌دهد که نسبت به روش‌های تجزیه سنتی مانند PCA یا SVD قابلیت تفسیر بیشتری دارد [۱٦]. دو ماتریس ایجادشده پس از تجزیه با عناوین ماتریس پایه و ماتریس مؤلفه نامیده می‌شوند [۱۷]. بازنمایی با ابعاد پایین همراه با ویژگی نامنفی بودن، قابلیت تفسیر را فراهم می‌کند، چراکه بازنمایی آموخته‌شده با روش ادراک انسان سازگارتر است و این توسط شواهد روان‌شناختی و فیزیولوژیکی ثابت‌شده است [۱۸]. به‌واسطه این ویژگی‌ها، NMF به یک تکنیک موردعلاقه در کاهش ابعاد در طیف گسترده‌ای از کاربردها ازجمله خوشه‌بندی [۱۹]، پیش‌بینی پیوند [۲۰]، انتخاب ویژگی [۲۱]، تکمیل ماتریس [۲۲] و جداسازی اختلاط در تصاویر فراطیفی [۲۳] در پارادایم‌های یادگیری بدون نظارت [۲۱]، خود نظارتی [۲٤] و نیمه نظارتی [۲٥] در نظر گرفته می‌شود.

مطالعات پیشین ارتباطات قوی بین NMF و الگوریتم‌های خوشه‌بندی کلاسیک را نشان داده‌اند و اثبات کرده‌اند که NMF تحت شرایط خاصی معادل با K–میانگین و خوشه‌بندی طیفی است [۱۰]. به‌ویژه، NMF استاندارد با معیار نرم فروبنیوس، ازنظر ریاضی معادل خوشه‌بندی با استفاده از K–میانگین است، که توضیح روشنی برای عملکرد رقابتی آن در یادگیری بدون نظارت ارائه می‌دهد. با الهام از این مطالعات، گونه‌های متعددی از NMF برای ارتقای قابلیت خوشه‌بندی آن پیشنهاد شده‌اند. به‌عنوان‌مثال، NMF متعامد [۲٦] محدودیت‌های تعامد را اعمال می‌کند تا افزونگی نمایش کاهش یابد و اجزای متمایزتر برجسته شوند و نمایه‌هایی ایجاد کند که به‌طور طبیعی با ساختار خوشه‌ها هماهنگ‌تر باشند. این‌گونه همچنین ارتباط نظری بین تجزیه مبتنی بر NMF و خوشه‌بندی مبتنی بر نمونه‌ها را تقویت می‌کند. توسعه‌های اضافی – مانند NMF خلوت [۲۷]، NMF با تنظیم‌گر گراف [۲۸] و مدل‌های سه‌عاملی مانند NMTF متعامد [۲۹]– با بهبود کیفیت نمایش، دربرگیری هندسه خمینه، یا افزایش انعطاف‌پذیری مدل‌سازی، چارچوب NMF را غنی‌تر می‌کنند و عموماً به‌عنوان تکمیل‌کننده‌های روش اصلی عمل می‌کنند. درمجموع، ترکیب خوشه‌بندی و تجزیه ماتریسی نامنفی به پژوهشگران امکان می‌دهد تا ساختارهای پنهان و الگوهای پیچیده در داده‌های چندبعدی و چندنمایی را با دقت و قابلیت تفسیر بالاتر استخراج کنند، که این امر اهمیت آن را در یادگیری بدون نظارت و کاربردهای عملی بیش‌ازپیش برجسته می‌کند.

یکی از محدودیت‌های اصلی مدل استاندارد NMF، آسیب‌پذیری آن در برابر نوفه و داده‌های پرت است. ازآنجاکه NMF کلاسیک خطای بازسازی را با استفاده از نرم فروبنیوس مربعی اندازه‌گیری می‌کند، حتی تعداد کمی داده پرت می‌تواند تابع هدف را تحت سلطه قرار داده و تجزیه را مختل کند. این حساسیت ناشی از فرض نوفه گاوسی ضمنی در تابع زیان L_2 است؛ بنابراین هنگامی‌که

داده‌ها شامل نوفه با توزیع غیرگاوسی (به‌ویژه توزیع‌های دم‌سنگین مانند توزیع لاپلاس یا کوشی) باشند یا نمونه‌های پرت بزرگ در داده وجود داشته باشند، مجازات مربعی انحرافات شدید را تشدید کرده و منجر به نمایه‌های ضعیف می‌شود.

برای رفع این ضعف، مدل‌های مختلف NMF استوار (Robust NMF) نرم فروبنیوس را با معیارهای استوارتر جایگزین کرده‌اند. حمزه و برادی [۳۰] مجازات L_1+L_2 را پیشنهاد دادند که مقاومت مدل را نسبت به داده‌های پرت افزایش می‌دهد. در [۳۱] NMF مبتنی بر L_1 برای مقابله با نوفه لاپلاسی ارائه‌شده است. کونگ و همکاران [۳۲] مدل پرکاربرد L_2,1 را توسعه دادند که از مربعی کردن باقی‌مانده‌های نمونه‌ای اجتناب کرده و بدین ترتیب اثر داده‌های پرت را محدود می‌کند. در ادامه همین رویکرد، لی و همکاران [۳۳] مقاومت مدل را با استفاده از مجازات غیرمحدب L_(۲,۰٫۵□) بیش‌تر بهبود دادند. رویکردهای دیگر بر اساس مفاهیم احتمالاتی یا نظریه اطلاعات برای افزایش مقاومت طراحی‌شده‌اند، ازجمله Cauchy-NMF [۳٤] که از توزیع دم‌سنگین کوشی بهره می‌برد، و مدل‌های مبتنی بر همسانی محلی مانند CIM-NMF [۳۵] که با معیارهای شباهت موضعی، انحرافات بزرگ را سرکوب می‌کنند. دسته دیگری از کارها هم اثر نمونه‌های آلوده را با محدود کردن خطاهای بزرگ کاهش می‌دهد، که ازجمله مهم‌ترین کارها می‌توان به Capped norm-NMF [۳٦] و Truncated Cauchy-NMF [۳۷] اشاره کرد؛ این‌گونه‌های استوار NMF به‌طور جمعی حساسیت ذاتی فرموله‌سازی نرم فروبنیوس کلاسیک را کاهش داده و عملکرد بهتری در حضور داده‌های پرت و نوفه غیرگاوسی فراهم می‌کنند. بااین‌حال، بسیاری از روش‌های استوار معمولاً فرض می‌کنند که مدل نوفه ثابت است و ممکن است در مواجهه با داده‌هایی با توزیع نوفه نامعلوم یا ناهمگن دچار مشکل شوند. این چالش انگیزه‌ای برای استفاده از بهینه‌سازی استوار توزیعی (DRO) ایجاد می‌کند، که به دنبال جواب‌هایی است که در بدترین حالت توزیع ممکن، درون یک مجموعه ابهام شامل توزیع‌های از پیش تعیین‌شده، عملکرد مناسبی داشته باشند [۳۸]. برخلاف بهینه‌سازی استوار کلاسیک (RO) که تنها اثر داده‌های پرت را با تغییر تابع زیان معمولی محدود می‌کند، DRO به‌صورت احتمالاتی عدم قطعیت داده‌ها را مدل می‌کند و تلاش می‌کند نه تنها برای یک نمونه مسئله مشخص، بلکه هم‌زمان در طیف گسترده‌ای از مسائل که هرکدام با توزیعی درمجموعه ابهام Ω تعیین می‌شوند، عملکرد مطلوب داشته باشد. به‌عبارت دیگر، در DRO فرض بر آن است که توزیع دقیق نوفه ناشناخته است، اما مجموعه‌ای از توزیع‌های مختلف در اختیار داریم و از قبل می‌دانیم که نوفه موجود در داده، یکی از توزیع‌های موجود در این مجموعه و یا ترکیبی از آن‌ها است؛ بنابراین می‌توان گفت، DRO فرآیند کمینه‌سازی تابع زیان مورد انتظار در بدترین حالت درمجموعه ابهام احتمالاتی است. این رویکرد منجر به جواب‌های استوارتر می‌شود و در سال‌های اخیر به دلیل تفسیر احتمالاتی داده‌های نامطمئن، قابلیت حل‌پذیری و عملکرد قابل‌توجه در نمونه‌های عددی، به حوزه‌ای فعال در پژوهش تبدیل‌شده است.

باوجود پتانسیل قابل‌توجه، یادگیری استوار توزیعی در زمینه NMF همچنان بسیار کم بررسی‌شده است و فضای گسترده‌ای برای تحقیقات بیشتر در این حوزه وجود دارد. به‌عنوان‌مثال، DR-NMF [۳۹] توزیع‌های نوفه مختلف را در نظر گرفته و تلاش می‌کند دقت و مقاومت تجزیه ماتریس نامنفی را در دامنه‌های گوناگون افزایش دهد. مدل‌های نوآورانه دیگر شامل NMF استوار توزیعی نمونه‌پایه (iDRNMF) [٤۰] و DRNMF-SP [٤۱] هستند. در iDRNMF، تابع هدف به‌صورت یک مسئله چندهدفه فرموله می‌شود، که در آن چندین تابع زیان ترکیب‌شده و این امکان را فراهم می‌کند که مدل نسبت به توزیع‌های نوفه متنوع استوار باشد و درعین‌حال ازنظر محاسباتی کارآمد باقی بماند. در DRNMF-SP با ترکیب مدل استوار توزیعی با یادگیری خودتکامل (Self-paced)، مدل به‌گونه‌ای آموزش می‌بیند که نمونه‌ها به‌تدریج بر اساس میزان دشواری (و از آسان به سخت) وارد فرآیند آموزش شوند؛ به‌این‌ترتیب، به‌طور طبیعی اثر نمونه‌های شدیداً آلوده (نوفه و داده‌های پرت) را کاهش داده و مشارکت توابع زیان مختلف را متوازن می‌کند.

تجزیه ماتریسی نامنفی به‌عنوان یکی از رویکردهای بنیادی برای نمایش جزءمحور و خوشه‌بندی داده‌های نامنفی، در حضور

نوفه‌های پیچیده و ساختارهای هندسی غیرخطی با چالش‌های جدی مواجه است. در بسیاری از کاربردهای واقعی، نوفه داده نه‌تنها دارای توزیع مشخص و ساده نیست، بلکه اغلب ترکیبی از توزیع‌های مختلف با دُم‌های سبک و سنگین است و در سطح نمونه‌ها (ستون‌ها) بروز می‌کند. این در حالی است که بخش قابل‌توجهی از مدل‌های NMF استوار، کیفیت تقریب را از طریق توابع زیان درایه‌محور و بر پایه یک توزیع نوفه مفروض می‌سنجند و درنتیجه ابهام آماری مانند مواجهه با سایر انواع نوفه و یا ترکیب چند نوع نوفه، عملکرد مطلوبی ندارند. از سوی دیگر، مدل‌های گراف منظم‌شده موجود، هرچند با هدف حفظ ساختار خمینه داده طراحی‌شده‌اند، این هدف را از طریق یک گراف k–نزدیک‌ترین همسایه با وزن‌های شباهت ثابت دنبال می‌کنند که به‌صورت ابتدایی ساخته‌شده و در طول فرایند تجزیه ثابت باقی می‌ماند. این وابستگی شدید به گراف اولیه کم‌کیفیت و فقدان به‌روزرسانی تطبیقیِ گراف، سبب می‌شود که ساختار محلی واقعی داده به‌درستی بازتاب نیابد و عملکرد خوشه‌بندی به‌طورجدی محدود شود.

افزون بر این، بیشتر رویکردهای استوار موجود یا بر جنبه «استوار بودن توزیعی» تمرکز کرده و ساختار هندسی داده را به‌طور صریح مدل نمی‌کنند، یا بر «یادگیری گراف» تأکید دارند و هنوز از یک چارچوب چندهدفه استوار توزیعی بهره نمی‌برند که بتواند به‌طور نظام‌مند با مجموعه‌ای از توابع هدف متناظر با توزیع‌های مختلف نوفه کار کند. درنتیجه، هنوز مدلی که بتواند (۱) نوفه‌های نامعلوم و بالقوه مخلوط را در سطح نمونه و در قالب یک چارچوب چندهدفه استوار توزیعی مدل کند، (۲) به‌طور هم‌زمان یک گراف مشابهت تطبیقی و دوتصادفی باکیفیت را از روی گراف اولیه کم‌کیفیت بیاموزد و (۳) همه این مؤلفه‌ها را در یک طرح بهینه‌سازی یکپارچه و با هزینه محاسباتی قابل‌مقایسه با NMF کلاسیک حل کند، به‌طور کامل شکل نگرفته است. این شکاف‌ها ضرورت طراحی مدلی را برجسته می‌کند که هم نسبت به طیف گسترده‌ای از نوفه‌های توزیع‌دار و ترکیبی استوار باشد و هم بتواند ساختار خمینه داده نوفه‌ای را به‌گونه‌ای تطبیقی و مبتنی بر گرافی باکیفیت در خود بازتاب دهد.

با توجه به مباحث فوق، چالش موجود را می‌توان در دو سطح متمایز اما مرتبط بیان کرد. سطح نخست، عدم قطعیت آماری مربوط به نوع و الگوی نوفه است. در داده‌های واقعی، معمولاً مشخص نیست که خطای بازسازی از یک توزیع مشخص، مانند گاوسی یا لاپلاس، پیروی می‌کند و حتی ممکن است چند الگوی نوفه‌ای به‌صورت هم‌زمان در داده وجود داشته باشند. در این حالت، انتخاب یک تابع زیان ثابت می‌تواند باعث شود مدل نسبت به بخشی از داده‌ها عملکرد مناسبی نداشته باشد. چارچوب iDRNMF این مسئله را با استفاده از یادگیری استوار توزیعی و وزن‌دهی تطبیقی توابع زیان موردتوجه قرار داده است؛ بااین‌حال، تمرکز آن بر جنبه آماری خطای بازسازی است.

سطح دوم، عدم قطعیت ساختاری مربوط به روابط میان نمونه‌ها و کیفیت گراف مورداستفاده برای نمایش ساختار محلی داده است. درروش‌های گراف منظم‌شده، گراف اولیه معمولاً با استفاده از روش‌هایی مانند k–نزدیک‌ترین همسایه ساخته می‌شود و سپس در فرآیند تجزیه ثابت باقی می‌ماند. درحالی‌که این گراف اولیه ممکن است به دلیل نوفه، داده‌های پرت یا پیچیدگی ساختار داده، روابط واقعی میان نمونه‌ها را به‌طور کامل نشان ندهد. بنابراین، حتی اگر خطای بازسازی به‌صورت استوار مدل شود، یک ساختار گرافی نامناسب می‌تواند کیفیت نمایش نهفته و درنتیجه کیفیت خوشه‌بندی را محدود کند.

از این دیدگاه، مقاومت در برابر نوفه و حفظ ساختار هندسی داده دو مسئله مستقل اما مکمل هستند. مقاومت توزیعی مشخص می‌کند که مدل در برابر عدم‌قطعیت آماری خطای بازسازی چگونه عمل کند، درحالی‌که یادگیری گراف مشخص می‌کند که روابط میان نمونه‌ها چگونه در فضای نهفته حفظ و اصلاح شوند. بر این اساس، مسئله اصلی این پژوهش صرفاً ترکیب چند تابع زیان نیست؛ بلکه طراحی یک چارچوب یکپارچه است که بتواند این دو سطح از عدم‌قطعیت را به‌صورت هم‌زمان در فرآیند یادگیری NMF مدیریت کند.

در این مقاله یک چارچوب جدید با عنوان تجزیه ماتریسی نامنفی استوار توزیعی مبتنی بر منظم‌ساز گراف دوتصادفی پیشنهاد می‌شود که به‌طور یکپارچه، مدل‌سازی استوار توزیعی نمونه‌محور و

یادگیری تطبیقی گراف را در یک مدل واحد ترکیب می‌کند. در این چارچوب، تابع زیان کلی بر پایه ترکیب وزنی پویا از چندین تابع هدف متناظر با توزیع‌های مختلف نوفه تعریف‌شده و مسئله بهینه‌سازی در قالب یک فرم حداقل–حداکثر بر روی مجموعه‌ای از توزیع‌های ابهام احتمالاتی فرمول‌بندی می‌شود تا بیشترین زیان مورد انتظار به‌صورت بدبینانه کمینه گردد. برای جلوگیری از غلبه یک تابع هدف با مقیاس بزرگ‌تر و تضمین مشارکت منصفانه همه اهداف، یک طرح نرمال‌سازی سیستماتیک برای توابع هدف به‌کار گرفته می‌شود که پایداری عددی و تفسیرپذیری ترکیب خطی اهداف را تضمین می‌کند. هم‌زمان، یک منظم‌ساز گراف دوتصادفی طراحی می‌شود که با یادگیری یک ماتریس شباهت نامنفی، متقارن و با مجموع سطری واحد، گراف اولیه کم‌کیفیت مبتنی بر k – نزدیک‌ترین همسایه را به گرافی غنی‌تر، تفکیک‌پذیرتر و سازگار با بازنمایی کم‌نوفه ارتقا می‌دهد.

نکته مهم آن است که وزن‌های مربوط به توابع زیان، پارامترهای قابل تنظیم دستی یا ابرپارامترهای ثابت مدل نیستند. این وزن‌ها به‌عنوان متغیرهای تصمیم در مسئله حداقل–حداکثر تعریف می‌شوند و در طول فرآیند بهینه‌سازی به‌صورت خودکار به‌روزرسانی می‌گردند. بنابراین، مدل نیازمند انتخاب دستی میزان اهمیت هر تابع زیان نیست؛ بلکه میزان مشارکت هر تابع زیان بر اساس رفتار خطای بازسازی و چارچوب استوار توزیعی تعیین می‌شود.

نوآوری‌های اصلی این پژوهش را می‌توان به‌صورت زیر خلاصه کرد:

۱) در این پژوهش، چارچوب استوار توزیعی نمونه‌محور iDRNMF با یک سازوکار یادگیری تطبیقی ساختار محلی ترکیب‌شده است. در این چارچوب، جمله استوار توزیعی مسئول مقابله با عدم‌قطعیت آماری ناشی از ناشناخته بودن الگوی نوفه است، درحالی‌که سازوکار گراف دوتصادفی، عدم‌قطعیت موجود در ساختار روابط میان نمونه‌ها را موردتوجه قرار می‌دهد. بدین ترتیب، مدل پیشنهادی به‌طور هم‌زمان یک نمایش استوار در برابر نوفه و یک ساختار نهفته منسجم‌تر را یاد می‌گیرد.

۲) برخلاف روش‌هایی که گراف اولیه را ثابت در نظر می‌گیرند، درروش پیشنهادی یک ماتریس شباهت نامنفی، متقارن و دوتصادفی به‌صورت درونزا از گراف اولیه یاد گرفته می‌شود. این امکان اصلاح روابط اولیه نمونه‌ها را در طول فرآیند یادگیری فراهم کرده و وابستگی مدل به کیفیت گراف اولیه را کاهش می‌دهد.

۳) دو سازوکار فوق در یک تابع هدف واحد ادغام‌شده‌اند و متغیرهای تجزیه، وزن‌های استوار توزیعی و ماتریس شباهت به‌صورت تکراری و هماهنگ به‌روزرسانی می‌شوند. همچنین، نرمال‌سازی توابع زیان و استفاده از وزن‌دهی مجدد تکراری امکان می‌دهد که تفاوت مقیاس توابع زیان و اثر نمونه‌های دارای خطای بزرگ کنترل شود.

٤) روش پیشنهادی بر روی ۱۲ مجموعه‌داده و در شرایط بدون نوفه و نوفه‌ای ارزیابی‌شده و علاوه بر مقایسه با روش‌های مختلف، با استفاده از مطالعه حذف مؤلفه‌ها نقش مستقل و مکمل وزن‌دهی تطبیقی و منظم‌ساز گراف دوتصادفی نیز بررسی‌شده است.

۲- پیش‌زمینه‌ها

در این بخش، مروری بر مقدمات کلیدی، ازجمله NMF استاندارد، $L_{2,1}-NMF$ و $Cauchy-NMF$ ارائه می‌شود. علاوه بر این، الگوریتم کمترین مربعات وزن‌دهی‌شده مجدد تکراری که در الگوریتم پیشنهادی استفاده می‌شود، معرفی می‌شود.

۲-۱ تجزیه ماتریسی نامنفی استاندارد

فرض کنید X ماتریس داده‌های ورودی باشد که حاوی عناصر نامنفی بوده و با X=[x_1,x_2,…,x_n]∈R^(m×n) نشان داده شود، که در آن m تعداد ویژگی‌ها و n تعداد نمونه‌ها است. هر بردار ستونی x_i یک نمونه داده نامنفی با m بعد را نشان می‌دهد. هدف NMF یافتن دو ماتریس نامنفی H∈R^(r×n) و W∈R^(m×r) است به‌گونه‌ای که بتوان ماتریس داده را با دقت مناسبی به‌صورت X≈WH بازسازی کرد. این هدف از طریق تابع زیر دنبال می‌شود:

$$\min_{W,H}\sum_{i=1}^{n}\Xi(x_i,Wh_i),s.t.\ \ W,H\geq 0 \qquad (۱)$$

در این رابطه، مطابق با تابع سنجش Ξ، هر نمونه x_i به‌عنوان ترکیبی خطی از پایه‌بردارهای موجود در W با ضرایبی که توسط بردار h_i تعیین می‌شود، بازسازی می‌گردد. مدل پایه NMF از فاصله خطای مربعی برای اندازه‌گیری اختلاف بین X و WH استفاده می‌کند و تابع هدف آن به‌صورت زیر تعریف می‌شود:

$$\min_{W,H} ||X - WH||_F^2 = \sum_{i=1}^{n} ||x_i - Wh_i||^2, \quad s.t. W, H \geq 0 \qquad (۲)$$

در اینجا، $||x_i - Wh_i||$ خطای بازسازی مربوط به نمونه i–ام را نشان می‌دهد و $||.||_F$ نرم فروبنیوس است. اگرچه مدل (۲) نسبت به هر یک از متغیرهای W و H به‌صورت مستقل محدب است، اما هنگامی‌که هر دو متغیر هم‌زمان در نظر گرفته شوند، خاصیت تحدب از بین می‌رود. از این‌رو، دستیابی به جواب بهینه سراسری امکان‌پذیر نیست و تنها می‌توان با بهره‌گیری از روش‌های بهینه‌سازی، به یک جواب بهینه محلی دست‌یافت.

شایان ذکر است که اکثر الگوریتم‌های بهینه‌سازی برای حل NMF ماهیت تکراری داشته و از این واقعیت بهره می‌گیرند که در صورت ثابت نگه‌داشتن یکی از ماتریس‌ها، مسئله به یک مسئله محدب کمترین مربعات نامنفی (NNLS) تبدیل می‌شود. به‌طور خاص، در هر تکرار یکی از دو عامل ثابت نگه‌داشته شده و عامل دیگر به‌گونه‌ای به‌روز می‌شود که مقدار تابع هدف کاهش یابد. مشهورترین و رایج‌ترین روش بهینه‌سازی در این زمینه، استفاده از قواعد به‌روزرسانی ضربی (MUR) است [۱۳]:

$$W \leftarrow W \odot \frac{XH^T}{WHH^T} \qquad (۳)$$

$$H \leftarrow H \odot \frac{W^T X}{W^T WH} \qquad (۴)$$

که در آن $\odot$ نشان‌دهنده ضرب عنصر به عنصر است.

۲-۲ الگوریتم کمترین مربعات وزن‌دهی‌شده مجدد تکراری

الگوریتم کمترین مربعات وزن‌دهی‌شده مجدد تکراری (IRLS) یک روش قدرتمند است که به‌طور گسترده برای بهینه‌سازی مدل‌های استوار به‌کار می‌رود. اگر خطای بازسازی نمونه i–ام با e_i نشان داده شود، مسائل بازسازی را می‌توان به‌صورت تابع هدف کلی زیر بیان کرد:

$$\min \sum_{i=1}^{n} \Xi(e_i) \qquad (۵)$$

که در آن تابع زیان $\Xi(e_i)$ باید با افزایش مقدار $|e_i|$ صعودی باشد. فرض کنید مسئله (۵) شامل m پارامتر با مقادیر نامعلوم باشد که با بردار $\boldsymbol{v} = [v_1, v_2, \dots, v_m]^T$ نمایش داده می‌شود. برای دستیابی به پاسخ بهینه، لازم است مشتق نسبت به این متغیرها صفر شود:

$$\sum_i \omega(e_i) \frac{\partial e_i}{\partial v_j} = 0, \quad j = 1,2,\dots,m \qquad (۶)$$

که در آن $\omega(e_i) = \frac{d\Xi(e_i)}{de_i}$ تابع تأثیر نام دارد. این رابطه را می‌توان به‌صورت زیر بازنویسی کرد:

$$\sum_i \psi(e_i) e_i \frac{\partial e_i}{\partial v_j} = 0, \quad j = 1,2,\dots,m \qquad (۷)$$

در اینجا $\psi(e_i)$ تابع وزن بوده و به این صورت تعریف می‌شود:

$$\psi(e) = \frac{\omega(e)}{e} = \frac{\Xi'(e)}{e} \qquad (۸)$$

روابط بالا نشان می‌دهد که معادله (۷) معادل حل مسئله بهینه‌سازی وزن‌دهی‌شده مجدد زیر است:

$$min \sum_i \psi(e_i^{[t-1]}) e_i^2 \qquad (۹)$$

برای حل مسئله (۹) در هر تکرار ابتدا وزن $\psi(e_i^{[t-1]})$ ثابت فرض شده و مقدار بهینه مسئله محاسبه می‌شود، سپس براساس خطای بازسازی جدید $e_i^{[t]}$، وزن‌ها مجدداً محاسبه و در تکرار بعد استفاده می‌گردند.

در [۴۲] ارتباط روش وزن‌دهی‌شده مجدد تکراری و NMF، مستقل از نوع تابع زیان، نشان داده‌شده است. در این چارچوب، تابع زیان NMF و خطای بازسازی ⟦|| ⟦x_i-Wh⟧ _i||⟧_ به‌ترتیب معادل Ξ(.) و e_i در رابطه (۵) هستند. بنابراین، رابطه (۱) را می‌توان به‌صورت زیر بازنویسی کرد:

$$\min_{W,H} \sum_{i=1}^{n} d_i ||x_i - Wh_i||^2, \quad s.t. \quad W, H \geq 0 \qquad (۱۰)$$

که در آن $d_i = \psi(||x_i - Wh_i||^{[t-1]})$ ضریب وزن نمونه i در هر تکرار است. با تشکیل ماتریس قطری D طوری که $D_{ii} = d_i$، قواعد به‌روزرسانی کلاسیک NMF در فرم وزنی به‌صورت زیر درمی‌آیند:

$$W \leftarrow W \odot \frac{XDH^T}{WHDH^T} \quad (۱۱)$$

$$H \leftarrow H \odot \frac{W^T XD}{W^T WHD} \quad (۱۲)$$

این بازنویسی نشان می‌دهد که الگوریتم IRLS می‌تواند فرآیند بهینه‌سازی NMF با انواع توابع زیان را برای مدل‌سازی استوار در برابر نمونه‌های پرت تقویت کند. مراحل بیان‌شده در Algorithm 1 ارائه‌شده است.

Algorithm 1. Iteratively Reweighted Least Squares (IRLS) for Robust NMF

Input: data matrix X, latent factor r, loss function $\Xi(\cdot)$;

Output: basis matrix W, representation matrix H, and weight matrix D;

1: Initialize $W \geq 0$ and $H \geq 0$;
2: **while** convergence not reached **do**
3: **for** i=1 to n:
4: Compute the reconstruction errore_i;
5: Compute the $\psi(e_i)$ according to Eq. (8);
6: Compute the sample weight d_i based on $\psi(e_i)$;
7: **end for;**
8: Construct the weight matrix D;
8: Update W according to Eq. (11);
9: Update H according to Eq. (12);
10: **end while**;
11: **return** W, H, and D.

$L_{2,1}$ – NMF ۳-۲

ازآنجاکه در NMF پایه، از خطای بازسازیِ مربعی برای تعریف تابع زیان استفاده می‌شود، این روش در حضور نوفه‌های بزرگ و داده‌های پرت، به‌شدت تحت تأثیر قرار می‌گیرد و عملکرد آن تضعیف می‌شود. برای افزایش مقاومت NMF در برابر نوفه و نقاط پرت، کونگ و همکاران [۳۲] روش L_2,1-NMF را پیشنهاد کردند که در آن، به‌جای استفاده از نرم فروبنیوس، از نرم L_2,1 برای اندازه‌گیری خطاهای بازسازی استفاده می‌شود. تابع هدف این مدل به‌صورت زیر تعریف می‌شود:

$$\min_{W,H} ||X - WH||_{2,1} = \sum_{i=1}^{n} ||x_i - Wh_i||, s.t. W, H \geq 0 \quad (۱۳)$$

در رابطه بالا، نرم $L_{2,1}$ ماتریس $M \in \mathbb{R}^{m \times n}$ به‌صورت $||M||_{2,1} = \sum_i^n \sqrt{\sum_j^m M_{ji}^2}$ تعریف می‌شود.. با توجه به عدم استفاده از توان دوم در خطای بازسازی اولیه، تأثیر نمونه‌های پرت کاهش‌یافته و این مدل نسبت به NMF استاندارد استوارتر عمل می‌کند. افزون بر این، در [۳۲] یک الگوریتم به‌روزرسانی تکراری کارآمد برای حل L_2,1-NMF ارائه‌شده که در آن از روابط (۱۱) و (۱۲) استفاده‌شده و ماتریس D به‌صورت زیر تعریف می‌شود:

$$D_{ii} = \frac{1}{||x_i - Wh_i||} \quad (۱۴)$$

Cauchy - NMF ۴-۲

در NMF استاندارد، نوفه معمولاً با توزیع گاوسی مدل‌سازی می‌شود و در L_2,1-NMF از توزیع لاپلاس برای مدل‌سازی نوفه استفاده می‌گردد. در [۳۴] فرض بر این است که نوفه دارای توزیع کوشی (Cauchy) و به‌صورت زیر بیان می‌شود:

$$p(x_i|Wh_i) \sim \frac{1}{\gamma} \frac{1}{1 + \left(\frac{||x_i - Wh_i||}{\gamma}\right)^2} \quad (۱۵)$$

که در آن γ پارامتر مقیاس است؛ مطابق این مدل، زمانی که اختلاف میان مقدار مشاهده‌شده و مقدار بازسازی‌شده (یعنی $||x_i - Wh_i||$) بزرگ باشد، مقدار احتمال در مخرج کاهش چشمگیری خواهد داشت؛ درنتیجه، مدل به‌طور طبیعی در برابر نوفه‌های شدید و نمونه‌های پرت استوارتر می‌شود؛ بر این اساس، می‌توان با توجه به مدل نوفه ارائه‌شده در رابطه (۱۵)، Cauchy-NMF را به شکل زیر ارائه کرد:

$$\min_{W,H} ||X - WH||_{2,cau} = \sum_{i=1}^{n} \ln\left(||x_i - Wh_i||^2 + \gamma^2\right), \quad s.t. W, H \geq 0 \quad (۱۶)$$

با بهره‌گیری از الگوریتم وزن‌دهی‌شده مجدد تکراری، می‌توان قواعد به‌روزرسانی وزنی را مشابه روابط (۱۱) و (۱۲) به‌دست آورد، با این تفاوت که ماتریس وزن D در این مدل به‌صورت زیر تعریف می‌شود:

$$D_{ii} = \frac{1}{||x_i - Wh_i||^2 + \gamma^2} \quad (۱۷)$$

۳- الگوریتم پیشنهادی

در سال‌های گذشته، چندین مدل NMF استوار توسعه داده‌شده‌اند که عمدتاً با تابع هدف انتخاب‌شده –که کیفیت یک تقریب را از

طریق فاصله بین WH و X ارزیابی می‌کند- مشخص می‌شوند. انتخاب این تابع هدف معمولاً توسط مدل نوفه فرضی روی ماتریس داده X تعیین و برای این کار، کاربران به‌صورت دستی تابع هدف را به‌صورت موردی انتخاب می‌کنند. در این روش‌ها، انتخاب نامناسب تابع هدف ممکن است منجر به یک جواب شود که به‌طور قابل‌توجهی با پاسخ موردنظر متفاوت باشد. پرداختن به این چالش شامل طراحی یک مدل NMF است که در برابر انواع مختلف توزیع نوفه استوار باشد و برای کاربردهای مختلف مناسب باشد. در این مقاله مدل تجزیه ماتریسی DRNMF-bS برای مدیریت انواع نوفه با توزیع‌های متفاوت ارائه‌شده است.

در این پژوهش، هدف ارائه یک تابع زیان جدید یا جایگزینی ساده یک تابع زیان با تابعی دیگر نیست. چارچوب پیشنهادی بر پایه iDRNMF بناشده و محدودیت اصلی آن را هدف قرار می‌دهد. iDRNMF می‌تواند عدم‌قطعیت مربوط به توزیع نوفه را با یادگیری تطبیقی وزن توابع زیان مدیریت کند، اما روابط محلی میان نمونه‌ها را به‌صورت صریح در فرآیند یادگیری کنترل نمی‌کند. درنتیجه، در DRNMF-bS تلاش می‌شود دو نوع اطلاعات به‌صورت هم‌زمان مورداستفاده قرار گیرد: نخست، اطلاعات آماری خطای بازسازی از طریق جمله استوار توزیعی و وزن‌دهی تطبیقی توابع زیان؛ و دوم، اطلاعات ساختاری نمونه‌ها از طریق یادگیری یک گراف دوتصادفی. بنابراین، نوآوری اصلی مدل پیشنهادی در یادگیری مشترک یک نمایش استوار توزیعی و یک ساختار گرافی تطبیقی در یک چارچوب NMF واحد است.

۳-۱ تجزیه ماتریسی نامنفی استوار توزیعی

این بخش مدل تجزیه ماتریسی نامنفی استوار توزیعی (iDRNMF) را در یک چارچوب چندهدفه معرفی می‌کند؛ مدلی که هدف آن یادگیری یک نمایش استوار از داده‌ها در سطح نمونه است. مدل پیشنهادی برای مقابله با طیف گسترده‌ای از نوفه با توزیع‌های مختلف طراحی‌شده و چارچوب یکپارچه آن قادر است هر تابع هدفی را که در ادبیات بهینه‌سازی استوار توزیعی مطرح‌شده است، دربرگیرد. برخلاف مدل‌های عنصرمحور، مدل پیشنهادی نمونه‌محور، هر ستون از ماتریس داده X را به‌عنوان یک نمونه مستقل در نظر گرفته و محاسبات را بر این اساس انجام می‌دهد.

به‌طور مشخص، یک تابع زیان کلی $||.||_{2,\tau}$ تعریف‌شده است که در آن ابتدا نرم ۲ به‌صورت مستقل بر هر ستون از ماتریس X اعمال‌شده و سپس زیان مدنظر τ بر خروجی‌ها محاسبه می‌شود؛ فرض می‌شود که τ با یک توزیع احتمالی مرتبط است، به‌گونه‌ای که $||.||_{2,\tau}$ امکان مدل‌سازی یک توزیع خاص برای نوفه را فراهم کند.

ازآنجاکه مدل دقیق نوفه داده‌ها ناشناخته است، فرض می‌شود قانون احتمال نوفه به یکی از توزیع‌های پایه یا ترکیبی از آن‌ها تعلق دارد. در این چارچوب، Ω مجموعه شاخص‌های توزیع‌ها و توابع زیان متناظر با آن‌هاست. برای هر $\tau \in \Omega$، توزیع احتمال متناظر با آن با P_τ و تابع زیان متناظر با L_τ نشان داده می‌شود؛ به‌گونه‌ای که L_τ اثر انتخاب توزیع P_τ را بر سنجش خطای بازسازی بیان می‌کند.

بر این اساس، مجموعه ابهام احتمالاتی متناظر با Ω، یعنی P_Ω، به‌صورت مجموعه تمامی ترکیب‌های محدب توزیع‌های پایه موجود در Ω تعریف می‌شود:

$$P_\Omega = \{P_\lambda | P_\lambda = \sum_{\tau\in\Omega} \lambda_\tau P_\tau, \lambda_\tau \geq 0, \sum_{\tau\in\Omega} \lambda_\tau = 1\}$$

در این تعریف، P_Ω مجموعه توزیع‌های احتمالی قابل‌قبول برای مدل‌سازی عدم‌قطعیت توزیعی نوفه است و λ پارامتر ترکیب این توزیع‌ها را تعیین می‌کند. بنابراین، Ω مجموعه توزیع‌های پایه موردبررسی و P_Ω مجموعه ابهام احتمالاتی متناظر با آن‌هاست. با این تعریف، چارچوب پیشنهادی محدود به یک توزیع مشخص نیست، بلکه می‌تواند برای هر مجموعه متناهی از توزیع‌های پایه و توابع زیان متناظر با آن‌ها به‌کار گرفته شود.

در این پژوهش، برای ارزیابی چارچوب پیشنهادی، $\Omega = \{1,2,cau\}$ در نظر گرفته‌شده است که به‌ترتیب متناظر با توابع زیان لاپلاسی، گاوسی و کوشی هستند. انتخاب این سه توزیع صرفاً به‌عنوان یک مطالعه موردی انجام‌شده است. این انتخاب امکان بررسی رفتار مدل در برابر طیفی از شرایط نوفه‌ای را فراهم می‌کند؛ به‌طوری‌که توزیع گاوسی نماینده خطاهای با رفتار دم‌سبک و پراکندگی محدود، توزیع لاپلاسی نماینده خطاهایی با احتمال بیشتر برای انحراف‌های بزرگ و رفتار دنباله‌ای بینابینی، و توزیع کوشی نماینده خطاهای بسیار دم‌سنگین و انحراف‌های شدید است. درنتیجه، این سه توزیع در کنار یکدیگر طیفی متنوع از رفتارهای

نوفه‌ای، از خطاهای معمول تا انحراف‌های شدید و دم‌سنگین، را پوشش می‌دهند.

لازم به تأکید است که ساختار مدل پیشنهادی به این مجموعه سه‌عضوی خاص وابسته نیست. با تغییر یا گسترش اعضای Ω مجموعه ابهام متناظر نیز تغییر می‌کند و همان چارچوب بهینه‌سازی می‌تواند برای مجموعه جدید به‌کار گرفته شود؛ مشروط بر آنکه تابع زیان متناظر با توزیع موردنظر در چارچوب وزن‌دهی مجدد مورداستفاده قابل‌اعمال باشد.

در مدل پیشنهادی، یک مجموع وزنی پویا از توابع هدف مختلف در نظر گرفته می‌شود که وزن‌های این توابع نیز آموخته می‌شوند. بدین ترتیب، راه‌حل پیشنهادی نسبت به انواع مختلف توزیع‌های نوفه و ترکیب آن‌ها استوار بوده و از طریق بهینه‌سازی یک ترکیب وزنی از چندین تابع هدف قابل‌حل است. بنابراین، در قالب یک فرمول‌بندی حداقل–حداکثر، هدف کمینه‌سازی بیشترین زیانِ موردانتظار روی مجموعه ابهام احتمالاتی Ω به‌صورت زیر است:

$$\min_{W,H}\max_{\lambda}\sum_{\tau\in\Omega}\lambda_\tau||X-WH||_{2,\tau}\ ,\quad s.t.W,H,\lambda\geq 0,||\lambda||_1=1 \qquad (۱۸)$$

در رابطه بالا، λ پارامتر توزیع P_λ درمجموعه ابهام P_Ω است. ازآنجاکه زیان متناظر با یک توزیع ترکیبی نسبت به توزیع‌های پایه خطی است، زیان مورد انتظار تحت P_λ به‌صورت ترکیب وزنی زیان‌های متناظر با توزیع‌های پایه بیان می‌شود. از این‌رو، بیشینه‌سازی نسبت به λ تحت قیود $\lambda\geq 0$ و $||\lambda||_1=1$، معادل با جست‌وجوی بدترین توزیع درمجموعه ابهام P_Ω است. بنابراین، رابطه (۱۸) یک فرم پارامتری‌شده از مسئله بهینه‌سازی استوار توزیعی را ارائه می‌کند که در آن، پارامتر λ توزیع بدترین حالت را از میان توزیع‌های موجود درمجموعه ابهام تعیین می‌کند.

درنتیجه، وزن‌های λ ضرایب ثابت و از پیش تعیین‌شده برای ترکیب توابع زیان نیستند، بلکه در فرآیند بهینه‌سازی به‌صورت تطبیقی تعیین می‌شوند تا توزیعی که بیشترین زیان را ایجاد می‌کند، در مسئله حداقل–حداکثر موردتوجه قرار گیرد.

در این پژوهش، مجموعه شاخص‌های پایه به‌صورت $\Omega=\{1,2,cau\}$ در نظر گرفته می‌شود که به‌ترتیب متناظر با توابع زیان لاپلاسی، گاوسی و کوشی هستند. بنابراین، مجموعه ابهام احتمالاتی متناظر، یعنی P_Ω ، شامل توزیع‌های پایه مذکور و ترکیب‌های محدب آن‌هاست. بنابراین، رابطه (۱۸) را می‌توان به‌شکل زیر بازنویسی کرد:

$$\begin{aligned}\min_{W,H}\max_{\lambda}\big(&\lambda_1||X-WH||_{2,1}\\&+\lambda_2||X-WH||_{2,2}^2\\&+\lambda_{cau}||X-WH||_{2,cau}\big),\\&s.t.\quad W,H,\lambda\geq 0,||\lambda||_1=1\end{aligned} \qquad (۱۹)$$

هدف (۱۹) را می‌توان به فرم نمونه‌محور زیر بیان کرد:

$$\begin{aligned}\min_{W,H}\max_{\lambda}\lambda_1&\sum_{i=1}^{n}||x_i-Wh_i||\\&+\lambda_2\sum_{i=1}^{n}||x_i-Wh_i||^2\\&+\lambda_{cau}\sum_{i=1}^{n}\ln\left(||x_i-Wh_i||^2+\gamma^2\right),\\&s.t.\quad W,H,\lambda\geq 0,||\lambda||_1=1\end{aligned} \qquad (۲۰)$$

برای بهینه‌سازی این چارچوب چندهدفه، ضروری است که در هر تکرار، مقدار کل زیان در رابطه (۱۸) کاهش یابد و به‌طور ویژه، بیشترین تمرکز بر روی زیان با بزرگ‌ترین مقدار است؛ به دلیل وجود توان دوم در نرم فروبنیوس، روشن است که نرم $||\cdot||_{2,2}$ مقادیر بزرگ‌تری ایجاد می‌کند. این ویژگی مانع از آن می‌شود که مدل بتواند تا رسیدن به همگرایی، سایر توابع زیان را به‌صورت مؤثر کمینه کند؛ بنابراین، مقیاس‌بندی توابع هدف ضروری است تا ترکیب خطی حاصل معنادار شده و هر جمله سهمی هم‌اهمیت بیابد. این اصلاح به‌ویژه برای مدل iDRNMF حیاتی است زیرا امکان تولید راه‌حل‌هایی با خطای اندک در تمامی توابع هدف را فراهم کرده و از تمرکز یک‌جانبه بر یک تابع خاص جلوگیری می‌کند. نتیجه آن است که مدل نهایی ویژگی‌های برتر مدل‌های تک‌هدفه مختلف را در خود ادغام می‌کند.

برای مقیاس‌بندی، ابتدا برای هر τ یک راه‌حل (W_τ,H_τ) برای مسئله $\min_{W,H\geq 0}||X-WH||_{2,\tau}$ به‌دست می‌آوریم تا خطای همگراشده $\varepsilon^{(\tau)}$ حاصل شود. سپس هر تابع هدف با استفاده از این مقدار ثابت نرمال می‌شود. با این کار می‌توان رابطه (۲۰) را با جای‌گذاری توابع هدف نرمال‌شده به شکل زیر بازنویسی کرد:

$$\min_{W,H}\max_{\lambda}\lambda_1\epsilon_1\sum_{i=1}^{n}||x_i-Wh_i|| + \lambda_2\epsilon_2\sum_{i=1}^{n}||x_i-Wh_i||^2 + \lambda_{cau}\epsilon_{cau}\sum_{i=1}^{n}\ln\left(||x_i - Wh_i||^2+\gamma^2\right), \quad s.t. \quad W,H,\lambda\geq 0, ||\lambda||_1 = 1 \tag{۲۱}$$

که در آن برای سادگی، $\epsilon_\tau = \frac{1}{\varepsilon(\tau)}$ تعریف‌شده است. توجه کنید که در غیاب این نرمال‌سازی، در اغلب موارد، مسئله عملاً به کمینه‌سازی تابع هدفی منجر می‌شود که مقدار بزرگ‌تری تولید می‌کند.

در اینجا لازم است تأکید شود که ضرایب λ_τ ابرپارامترهایی نیستند که کاربر پیش از اجرای الگوریتم به‌صورت دستی تعیین کند. این ضرایب متغیرهای تصمیم مسئله حداقل–حداکثر هستند و تحت قید $||\lambda||_1 = 1$ به‌صورت خودکار یاد گرفته می‌شوند. درنتیجه، مجموع وزن‌ها برابر یک است و مدل می‌تواند در هر تکرار، میزان اهمیت نسبی توابع زیان متناظر با توزیع‌های مختلف را تغییر دهد. به بیان ساده، به‌جای آنکه کاربر از پیش مشخص کند هر تابع زیان چه سهمی در مدل داشته باشد، چارچوب استوار توزیعی این سهم‌ها را در فرآیند بهینه‌سازی تعیین می‌کند. بنابراین، استفاده از چند تابع زیان در این پژوهش معادل با یک ترکیب خطی ثابت و دستی نیست؛ بلکه بخشی از مسئله بهینه‌سازی است و وزن‌ها هم‌زمان با نمایش نهفته یاد گرفته می‌شوند.

ازآنجاکه بخش بیشینه‌سازی مسئله بر روی مجموعه وزن‌های نامنفی با مجموع واحد انجام می‌شود، وزن‌ها به‌صورت تطبیقی به سمت تابع زیانی هدایت می‌شوند که در وضعیت فعلی، ریسک بیشتری ایجاد می‌کند. بنابراین، مدل به‌جای انتخاب یک تابع زیان ثابت، در هر مرحله به بدترین مؤلفه موجود درمجموعه ابهام توجه بیشتری نشان می‌دهد. این ویژگی، اساس استوار بودن توزیعی مدل است: اگر نوع نوفه از قبل مشخص نباشد، مدل مجبور نیست یک تابع زیان را به‌صورت قطعی انتخاب کند، بلکه می‌تواند در طول فرآیند یادگیری اهمیت نسبی توابع زیان مختلف را تغییر دهد.

به‌طورکلی، ویژگی اصلی چارچوب پیشنهادی، وابستگی آن به یک مجموعه مشخص از توزیع‌های نوفه نیست، بلکه قابلیت آن در مدل‌سازی عدم‌قطعیت نسبت به مجموعه‌ای از توزیع‌های ممکن است. بنابراین، Ω را می‌توان متناسب با نوع داده و سناریوی نوفه‌ای موردنظر تعریف یا گسترش داد و مجموعه ابهام P_Ω را بر اساس توزیع‌های انتخاب‌شده تشکیل داد. در این پژوهش، ترکیب گاوسی، لاپلاسی و کوشی تنها به‌عنوان یک مجموعه نماینده برای پوشش رفتارهای دم‌سبک، بینابینی و بسیار دم‌سنگین انتخاب‌شده است.

۳-۲ iDRNMF وزن‌دهی مجدد تکراری

در بخش ۲–۲، الگوریتم بازوزن‌دهی تکراری ارائه و نشان داده شد که بر اساس آن می‌توان هر تابع هدفی را به یک مدل پایه NMF وزن‌دار تبدیل کرد. بنابراین، با بهره‌گیری از چارچوب بازوزن‌دهی معادله (۱۰)، فرمول چندهدفه (۱۸) به‌صورت زیر نوشته می‌شود:

$$\min_{W,H}\max_{\lambda}\sum_{i=1}^{n}d_i^{(\Omega)}||x_i-Wh_i||^2, \quad s.t. \quad W,H,\lambda\geq 0, ||\lambda||_1 = 1 \tag{۲۲}$$

که در آن، وزن نمونه $d_i^{(\Omega)}$ مطابق رابطه (۹) به‌صورت زیر محاسبه می‌شود:

$$d_i^{(\Omega)} = \sum_{\tau\in\Omega}\lambda_\tau\epsilon_\tau\psi_\tau(||x_i-Wh_i||^{[t-1]}) \tag{۲۳}$$

در واقع، در رابطه (۲۲) هر بخش از تابع هدف دارای یک وزن $d^{(\tau)}$ است؛ بنابراین می‌توان آن را به‌شکل زیر بازنویسی کرد:

$$\min_{W,H}\max_{\lambda}\sum_{\tau\in\Omega}\lambda_\tau\epsilon_\tau\sum_{i=1}^{n}d_i^{(\tau)}||x_i-Wh_i||^2, \quad s.t. \quad W,H,\lambda\geq 0, ||\lambda||_1 = 1 \tag{۲۴}$$

که در آن وزن نمونه برابر $d_i^{(\tau)} = \psi_\tau(||x_i-Wh_i||^{[t-1]})$ است. اگر $\Omega = \{1,2,cau\}$ آنگاه:

$$\min_{W,H}\max_{\lambda}\lambda_1\epsilon_1\sum_{i=1}^{n}d_i^{(1)}||x_i-Wh_i||^2 + \lambda_2\epsilon_2\sum_{i=1}^{n}d_i^{(2)}||x_i-Wh_i||^2 + \lambda_{cau}\epsilon_{cau}\sum_{i=1}^{n}d_i^{(cau)}||x_i-Wh_i||^2, \quad s.t.\quad W,H,\lambda\geq 0, ||\lambda||_1=1 \tag{۲۵}$$

که در آن به ترتیب ضرایب $d_i^{(1)}=\frac{1}{||x_i-Wh_i||}$، $d_i^{(2)}=1$ و $d_i^{(cau)}=\frac{1}{||x_i-Wh_i||^2+\gamma^2}$ است. درنهایت، می‌توان فرمول را به‌صورت تجمیع‌شده زیر بازنویسی کرد:

$$\min_{W,H}\max_{\lambda}\sum_{i=1}^{n}\left(\lambda_1\epsilon_1d_i^{(1)}+\lambda_2\epsilon_2+\lambda_{cau}\epsilon_{cau}d_i^{(cau)}\right)||x_i-Wh_i||^2 = \sum_{i=1}^{n}d_i^{(\Omega)}||x_i-Wh_i||^2,\quad s.t.\quad W,H,\lambda\geq 0,\quad ||\lambda||_1=1 \tag{۲۶}$$

که در آن $d_i^{(\Omega)}=\lambda_1\epsilon_1d_i^{(1)}+\lambda_2\epsilon_2+\lambda_{cau}\epsilon_{cau}d_i^{(cau)}$. شایان ذکر است که مجموعه Ω را می‌توان به‌دلخواه گسترش داد تا هر توزیع مطلوبی را دربرگیرد و از این فرمول‌بندی وزن‌دار یکپارچه برای مدیریت آن استفاده کرد.

۳-۳ منظم‌ساز گراف دو تصادفی

پس از مدیریت عدم‌قطعیت آماری مربوط به الگوی نوفه، مسئله دوم به ساختار روابط میان نمونه‌ها مربوط می‌شود. در NMF گراف منظم شده، کیفیت نمایش نهفته به کیفیت گراف شباهت وابسته است. اگر گراف اولیه روابط نادرست یا ناقصی میان نمونه‌ها ایجاد کند، این خطا می‌تواند از طریق منظم‌ساز گراف وارد فضای نهفته شود. ازاین‌رو، در این پژوهش گراف اولیه به‌عنوان یک ساختار کاملاً صحیح و ثابت فرض نمی‌شود؛ بلکه به‌عنوان یک تخمین اولیه از روابط نمونه‌ها در نظر گرفته‌شده و یک ماتریس شباهت اصلاح‌شده و دوتصادفی در طول فرآیند یادگیری از آن استخراج می‌شود. این بخش، سطح دوم عدم‌قطعیت موردتوجه در مدل پیشنهادی، یعنی عدم‌قطعیت ساختاری، را مدیریت می‌کند.

به دلیل وجود قیود نامنفی بر ماتریس‌های W و H مدل NMF قادر است یک نمایش مبتنی بر اجزاء از ماتریس داده X ارائه کند. ازاین‌رو، NMF برای داده‌های نامنفی (همچون اسناد و تصاویر چهره) تفسیر فیزیولوژیک و روان‌شناختی مناسب‌تری نسبت به روش‌های دیگر ارائه می‌دهد. بااین‌حال، NMF تنها ساختار کلی داده X را مدل می‌کند و به روابط محلی بین نمونه‌ها نمی‌پردازد. برای رفع این ضعف، چندین گونه گراف منظم شده از NMF ارائه‌شده‌اند.

یکی از نسخه‌های شاخص، GNMF [43] است که روابط محلی میان نمونه‌ها را با استفاده از گراف k-نزدیک‌ترین همسایه مدل‌سازی می‌کند. GNMF مسئله زیر را حل می‌کند:

$$\min_{W,H}||X-WH||_F^2+\gamma Tr(HL_SH^T),\quad s.t.W,H\geq 0 \tag{۲۷}$$

که در آن $L_S=U_S-S$ گراف لاپلاسین، S ماتریس شباهت و U_S ماتریس قطری درجه با عناصر $U_{ii}=\sum_j^n S_{ij}$ است. در رابطه بالا، جمله دوم تضمین می‌کند که نمونه‌هایی با شباهت بیشتر، فاصله کمتری در فضای پنهان داشته باشند و به‌این‌ترتیب، ویژگی‌های هندسی محلی داده در مدل لحاظ شود.

بااین‌حال، گراف ساخته‌شده توسط k-نزدیک‌ترین همسایه بسیار ساده و در بسیاری از موارد فاقد قدرت تفکیک‌پذیری کافی است و نمی‌تواند ساختار واقعی داده را به‌خوبی بازنمایی کند. علاوه‌براین، گراف ورودی در فرآیند تجزیه ثابت است و به‌روزرسانی نمی‌شود. در این مقاله، هدف ساخت یک گراف باکیفیت‌تر برای ارتقای عملکرد NMF است، به‌گونه‌ای که نتیجه خوشه‌بندی نیز بهبود یابد؛ بنابراین، با داشتن یک گراف اولیه کم‌کیفیت G، تلاش می‌شود ماتریس شباهت باکیفیت‌تر S بر اساس آن یاد گرفته شود.

مقدار هر درایه S میزان شباهت میان دو نمونه متناظر را نشان می‌دهد و مجموع هر سطر آن برابر یک است؛ این قید ماتریس S را به توزیع احتمال همسایگی هر نمونه بر سایر نمونه‌ها تبدیل می‌کند؛ این بدان معناست که $S1_n=1_n$ که 1_n بردار با مولفه‌های تمام یک است و بنابراین می‌توان نوشت $L_S=U_S-S=I-S$. همچنین باید میزان شباهت بین نقاط را بازتاب دهد، بنابراین لازم است نامنفی و متقارن باشد؛ یعنی $S=S^T$ و $S\geq 0$

ماتریسی با این ویژگی‌ها در ادبیات پژوهش به‌عنوان ماتریس دوتصادفی شناخته می‌شود [44]. به‌علاوه طبق نظریه گراف، هیچ رأسی نباید به خودش متصل باشد؛ بنابراین $diag(S) = 0$. با اعمال مجموعه قیود فوق، ماتریس S را می‌توان با حل مسئله زیر از گراف ورودی G یاد گرفت:

$$\min_{S} ||S - G||_F^2, s.t. S \geq 0, S = S^T, S\mathbf{1}_n = \mathbf{1}_n, diag(S) = 0 \quad (۲۸)$$

اکنون دو سطح عدم‌قطعیت در یک چارچوب واحد با یکدیگر ترکیب می‌شوند. بخش استوار توزیعی، عدم‌قطعیت آماری ناشی از ناشناخته بودن الگوی نوفه را از طریق چند تابع زیان و وزن‌دهی تطبیقی آن‌ها مدیریت می‌کند؛ در مقابل، منظم‌ساز گراف دوتصادفی، عدم‌قطعیت ساختاری ناشی از ناکامل یا نامناسب بودن گراف اولیه را با یادگیری یک ماتریس شباهت جدید کنترل می‌کند. بنابراین، دو مؤلفه مدل نقش یکسانی ندارند و صرفاً دو جمله مستقل به تابع هدف اضافه نشده‌اند: مؤلفه اول بر میزان اعتماد به خطای بازسازی در شرایط نوفه‌ای و مؤلفه دوم بر سازگاری روابط نمونه‌ها در فضای نهفته تمرکز دارد. هدف مدل پیشنهادی، یادگیری هم‌زمان این دو جنبه در یک فرآیند تجزیه واحد است. با ترکیب این دو مؤلفه، چارچوب تجزیه DRNMF-bS به شکل زیر پیشنهاد می‌شود:

$$\min_{W,H,S} \max_{\lambda} \sum_{\tau \in \Omega} \lambda_\tau ||X - WH||_{2,\tau} + \alpha Tr(HL_SH^T) + \beta ||S - G||_F^2, \quad s.t. W, H, \lambda, S \geq 0, ||\lambda||_1 = 1, S = S^T, S\mathbf{1}_n = \mathbf{1}_n, diag(S) = 0 \quad (۲۹)$$

که در آن α و β پارامترهای تنظیم هستند؛ α وزن جمله منظم‌ساز خمینه را کنترل می‌کند که با بهره‌گیری از ماتریس شباهت آموخته‌شده S، ساختار محلی داده‌ها را در فضای نهفته حفظ می‌کند، و β وزن جمله برازش را کنترل می‌کند که ماتریس شباهت آموخته‌شده S را به گراف اولیه G نزدیک نگه می‌دارد. رابطه (۲۹) نشان می‌دهد که مدل پیشنهادی تنها یک ترکیب ثابت از چند تابع زیان نیست. متغیرهای W و H نمایش داده را یاد می‌گیرند، λ میزان اهمیت نسبی توابع زیان را در چارچوب حداقل–حداکثر تعیین می‌کند، و S ساختار ارتباطی میان نمونه‌ها را اصلاح می‌کند. ازاین‌رو، سه جزء اصلی مدل به‌صورت مشترک با یکدیگر تعامل دارند: جمله استوار توزیعی برای مدیریت عدم‌قطعیت آماری، بازنمایی نامنفی برای استخراج ویژگی‌های نهفته، و یادگیری گراف دوتصادفی برای حفظ و اصلاح ساختار هندسی داده.

نقش وزن‌های تطبیقی در فرآیند یادگیری: نکته مهم در چارچوب پیشنهادی آن است که وزن‌های λ صرفاً ضرایبی برای تعیین سهم نهایی توابع زیان در تابع هدف نیستند، بلکه به‌صورت پویا و هم‌زمان با یادگیری عوامل تجزیه، در فرآیند شکل‌گیری نمایش نهفته نقش دارند. همان‌گونه که در رابطه (۲۳) بیان شد، وزن نمونه‌ها بر اساس ترکیب وزن‌دار توابع زیان و خطای بازسازی فعلی محاسبه می‌شوند. ازاین‌رو، تغییر وزن‌های λ موجب تغییر وزن نمونه‌ها و درنتیجه تغییر ماتریس D می‌شود. ازآنجاکه ماتریس D در روابط به‌روزرسانی W و H وارد می‌شود، مقدار λ در هر مرحله می‌تواند برجهت و میزان تغییر عوامل تجزیه در مراحل بعدی اثر بگذارد. در مقابل، تغییر W و H موجب تغییر خطاهای بازسازی‌شده و این تغییرات نیز در تعیین مقادیر جدید λ مؤثر هستند. بنابراین، یادگیری λ و یادگیری عوامل W و H دو فرآیند مستقل از یکدیگر نیستند، بلکه به‌صورت متقابل و تکرارشونده بر یکدیگر اثر می‌گذارند.

این ویژگی باعث می‌شود که فرآیند بهینه‌سازی مدل پیشنهادی ماهیتی مسیرمحور داشته باشد. به بیان دقیق‌تر، هر مقدار میانی λ تنها بیانگر وضعیت موقت وزن‌دهی توابع زیان نیست، بلکه با اثرگذاری بر وزن نمونه‌ها، وضعیت متغیرهای W و H در مرحله بعدی را نیز تعیین می‌کند. درنتیجه، دنباله مقادیر λ در طول تکرارهای الگوریتم می‌تواند مسیر حرکت W و H در فضای تصمیم را شکل دهد و متقابلاً وضعیت W و H نیز مسیر تغییرات بعدی λ را تعیین کند. از این منظر، متغیرهای W ، H، Sو λ را می‌توان اجزای یک فضای تصمیم مشترک در نظر گرفت که در آن یادگیری هر جزء به وضعیت سایر اجزا وابسته است.

اهمیت این ویژگی در مسئله حاضر، که نسبت به متغیرهای تجزیه غیرمحدب است، بیشتر نمایان می‌شود. در چنین مسائلی، هدف صرفاً یافتن یک مجموعه وزن نهایی برای ترکیب توابع زیان نیست؛

بلکه نحوه تعامل این وزن‌ها با عوامل تجزیه در طول فرآیند بهینه‌سازی نیز می‌تواند بر نقطه ایستای حاصل اثرگذار باشد. بنابراین، برتری سازوکار پیشنهادی صرفاً ناشی از در اختیار داشتن چند تابع زیان به‌طور هم‌زمان نیست، بلکه از یادگیری تطبیقی سهم این توابع در حین یادگیری نمایش نیز ناشی می‌شود. به‌عبارت‌دیگر، λ درروش پیشنهادی یک ابرپارامتر تنظیم‌شونده توسط کاربر نیست، بلکه بخشی از فرآیند یادگیری است که متناسب با وضعیت فعلی داده و خطای بازسازی، سهم توابع زیان را تعیین کرده و هم‌زمان بر مسیر بهینه‌سازی عوامل W و H اثر می‌گذارد.

۳-٤ بهینه‌سازی

ازآنجاکه تابع هزینه مدل پیشنهادی (۲۹) به‌صورت هم‌زمان نسبت به W و H محدب نیست، بهینه‌سازی مستقیم آن می‌تواند با چالش‌هایی روبه‌رو شود و احتمال رسیدن به جواب‌های غیربهینه وجود دارد. برای رفع این مشکل، مسئله اصلی را به چند زیرمسئله تقسیم می‌کنیم که هر یک، با ثابت نگه‌داشتن سایر متغیرها، نسبت به یکی از متغیرها محدب است. بر این اساس، متغیرها به‌صورت تکراری و یکی پس از دیگری به‌روزرسانی می‌شوند. در هر تکرار، با ثابت نگه‌داشتن سایر متغیرها، از روش گرادیان کاهشی برای به‌روزرسانی متغیر دیگر استفاده می‌کنیم تا فرایند بهینه‌سازی به‌تدریج به یک جواب مناسب همگرا شود.

قبل از استخراج قاعده به‌روزرسانی، لازم است شرایط بهینگی کاروش-کوهن-تاکر (KKT) برای مسئله‌ای با قید نامنفی بودن مرور شود. برای یک مسئله کلی به شکل $\min f(X)$ به‌شرط $X \geq 0$، اگر Θ ضریب لاگرانژ متناظر با قید $X \geq 0$ باشد، تابع لاگرانژی به‌صورت زیر تعریف می‌شود:

$$\mathcal{L}(X, \Theta) = f(X) - Tr(\Theta^T X)$$

شرایط KKT برای این مسئله عبارت‌اند از [45]:

$$\frac{\partial f}{\partial X} - \Theta = 0 \quad \text{(i)}$$
$$X \geq 0 \quad \text{(ii)}$$
$$\Theta \geq 0 \quad \text{(iii)}$$
$$\Theta \odot X = 0 \quad \text{(iv)}$$

به‌روزرسانی W: برای به‌دست‌آوردن فرمول به‌روزرسانی W با روش MUR، تابع هدف (۲۹) را می‌توان با استفاده از عملگر Trace بازنویسی کرد؛ سپس، ضرایب لاگرانژ $\boldsymbol{\Theta}$ و $\boldsymbol{\Phi}$ را تعریف می‌کنیم که به‌ترتیب نامنفی بودن W و H را تضمین می‌کنند:

$$\begin{aligned}\mathcal{L} = Tr(&X\mathcal{D}X^T - 2X\mathcal{D}H^TW^T \\ &+ WH\mathcal{D}H^TW^T) \\ &+ \alpha Tr(HL_SH^T) \\ &+ \beta Tr(SS^T - 2SG^T \\ &+ GG^T) - Tr(\boldsymbol{\Theta}^TW) \\ &- Tr(\boldsymbol{\Phi}^TH),\end{aligned} \quad (۳۰)$$

که در آن، $\Theta = [\theta_{ij}]$ و $\Phi = [\phi_{ij}]$ ماتریس ضرایب لاگرانژ برای قیود $w_{ij} \geq 0$ و $h_{ij} \geq 0$ هستند. با گرفتن مشتق جزئی نسبت به W داریم:

$$\frac{\partial \mathcal{L}}{\partial W} = -2X\mathcal{D}H^T + 2WH\mathcal{D}H^T - \Theta, \quad (۳۱)$$

با اعمال شرایط KKT بر مسئله (۲۹) نسبت به متغیر W، از شرط (i) و رابطه (۳۱) داریم:

$$\Theta = -2X\mathcal{D}H^T + 2WH\mathcal{D}H^T, \quad (۳۲)$$

با جای‌گذاری این رابطه در شرط مکملی (iv)، یعنی $W \odot \Theta = 0$، معادله زیر حاصل می‌شود:

$$W \odot -2X\mathcal{D}H^T + 2WH\mathcal{D}H^T = 0, \quad (۳۳)$$

با حل (۳۳) به روش MUR، قانون به‌روزرسانی زیر برای W حاصل می‌شود:

$$W \leftarrow W \odot \frac{X\mathcal{D}H^T}{WH\mathcal{D}H^T} \quad (۳٤)$$

که در آن ضرب و تقسیم به‌صورت مؤلفه به مؤلفه (هادامارد) انجام می‌شود.

به‌روزرسانی H: به همین ترتیب، برای استخراج قاعده به‌روزرسانی متغیر H، لازم است مشتق جزئی تابع لاگرانژی (۳۰) نسبت به H محاسبه شود. این تابع شامل دو بخش است: بخش خطای بازسازی موزون و بخش منظم‌ساز گراف. مشتق بخش بازسازی، مشابه با مسیر طی‌شده برای W، به‌صورت زیر به‌دست می‌آید:

$$\begin{aligned}&\frac{\partial[Tr(XDX^T - 2XDH^TW^T + WHDH^TW^T)]}{\partial H} \\ &= -2W^TXD \\ &+ 2W^TWHD,\end{aligned}$$

برای مشتق بخش منظم‌ساز گراف، از این اتحاد استاندارد جبر ماتریسی استفاده می‌شود که برای هر ماتریس متقارن M، مشتق $Tr(HMH^T)$ نسبت به H برابر است با $2HM$. ازآنجاکه ماتریس

لاپلاسین $L_S = U_S - S$ متقارن است (چون هم U_S و هم S متقارن هستند)، داریم:

$$\frac{\partial[\mathrm{Tr}(\mathrm{HL_SH^T})]}{\partial \mathrm{H}} = 2\mathrm{HL_S} = 2\mathrm{H(U_S - S)} = 2\mathrm{HU_S} - 2\mathrm{HS}$$

با ترکیب دو بخش فوق و با افزودن وزن α، مشتق کامل تابع لاگرانژی نسبت به H به‌صورت زیر حاصل می‌شود:

$$\frac{\partial \mathcal{L}}{\partial H} = -2W^T X\mathcal{D} + 2W^T WH\mathcal{D} + 2\alpha HU_S - 2\alpha HS - \Phi,$$

با برابر صفر قرار دادن این مشتق طبق شرط ایستایی KKT، ضریب لاگرانژ $\boldsymbol{\Phi}$ به دست می‌آید:

$$\Phi = -2W^T X\mathcal{D} + 2W^T WH\mathcal{D} + 2\alpha HU_S - 2\alpha HS$$

سپس با اعمال شرط مکملی $H \odot \Phi = 0$ داریم:

$$H \odot (-2W^T X\mathcal{D} + 2W^T WH\mathcal{D} + 2\alpha HU_S - 2\alpha HS) = 0$$

با بازآرایی این رابطه به شکل MUR قاعده به‌روزرسانی زیر حاصل می‌شود:

$$H \leftarrow H \odot \frac{W^T X\mathcal{D} + \alpha HS}{W^T WH\mathcal{D} + \alpha HU_S} \quad (۳۵)$$

به‌روزرسانی $\mathbf{S}$: با فرض ثابت بودن سایر متغیرها، S را بر اساس تابع هدف زیر می‌توان به دست آورد:

$$\min_S \alpha Tr(HL_S H^T) + \beta||S - G||_F^2, s.t. S \geq 0, S = S^T, S\mathbf{1}_n = \mathbf{1}_n, diag(S) = 0 \quad (۳۶)$$

با تعریف $M = G + \frac{\alpha}{2\beta} H^T H$ مسئله بالا را به‌صورت زیر می‌توان بازنویسی کرد:

$$\min_S ||S - M||_F^2, \quad s.t.\ S \geq 0, S = S^T, S\mathbf{1}_n = \mathbf{1}_n, diag(S) = 0 \quad (۳۷)$$

به‌جای حل مسئله (۳۷)، می‌توان آن را به دو زیرمسئله زیر تقسیم کرد:

$$\min_S ||S - M||_F^2, \quad s.t. S = S^T, S\mathbf{1}_n = \mathbf{1}_n \quad (۳۸)$$

$$\min_S ||S - M||_F^2, \quad s.t. S \geq 0, diag(S) = 0 \quad (۳۹)$$

سپس دو زیرمسئله بالا به ترتیب زیر حل می‌شوند: ۱) ابتدا S_1 از زیرمسئله (۳۸) به‌دست آمده و به‌عنوان M در زیرمسئله (۳۹) در نظر گرفته می‌شود؛ ۲) پس‌ازآن، S_2 از زیرمسئله (۳۹) حل‌شده و به‌عنوان M در زیرمسئله (۳۸) در نظر گرفته می‌شود؛ این کار تا رسیدن به همگرایی تکرار می‌شود. بر اساس [46] حل بهینه زیرمسئله (۳۸) برابر است با:

$$\mathrm{S_1} = \mathrm{M} + \frac{\mathrm{n} + 1_\mathrm{n}{}^\mathrm{T}\mathrm{M}1_\mathrm{n}}{\mathrm{n}^2} 1_\mathrm{n}1_\mathrm{n}{}^\mathrm{T} - \frac{1}{\mathrm{n}}\mathrm{M}1_\mathrm{n}1_\mathrm{n}{}^\mathrm{T} - \frac{1}{\mathrm{n}}1_\mathrm{n}1_\mathrm{n}{}^\mathrm{T}\mathrm{M} \quad (۴۰)$$

همچنین زیرمسئله (۳۹) به‌سادگی از رابطه زیر به دست می‌آید:

$$S_2 = \max(M, 0), diag(S_2) = 0 \quad (۴۱)$$

به‌روزرسانی وزن‌ها: ایده اصلی برای حل مسئله بهینه‌سازی (۲۹) آن است که در هر تکرار، بدترین حالت کمینه شود. بنابراین، باید بیشینه مقادیر میان توابع هدف را یافته و تلاش شود آن را در تکرار بعدی کاهش داد. به‌بیان‌دیگر، در هر تکرار، اگرچه تمامی توابع هدف کاهش می‌یابند، بزرگ‌ترین آن‌ها باید در تکرار بعدی از اهمیت بیشتری برخوردار باشد. به همین دلیل، از الگوریتم فرانک–وولف که بانام روش گرادیان شرطی نیز شناخته می‌شود استفاده‌شده است.

انتخاب این روش بر پایه سه ویژگی ساختاری مسئله استوار است. نخست، تابع هدف داخلی نسبت به λ خطی است و مجموعه شدنی آن، یعنی سیمپلکس احتمالاتی $\{\lambda \geq 0, \|\lambda\|1=1\}$، یک مجموعه محدب و فشرده است؛ برای این دسته از مسائل، حل زیرمسئله خطی‌شده فرانک–وولف در هر تکرار به یک عملیات ساده arg max روی رئوس سیمپلکس (یعنی انتخاب بردار one hot متناظر با بزرگ‌ترین مقدار تابع هدف) تقلیل می‌یابد و نیازی به حل زیرمسئله بهینه‌سازی اضافی ندارد. دوم، برخلاف روش‌های مبتنی بر گرادیان تصویری که در هر گام مستلزم تصویر کردن نقطه به‌دست‌آمده بر روی سیمپلکس (عملیاتی با نیاز به مرتب‌سازی مؤلفه‌ها با مرتبه پیچیدگی $O(|\Omega| \log|\Omega|)$) هستند، فرانک–وولف از این هزینه محاسباتی اضافه بی‌نیاز است (هزینه محاسباتی $O(|\Omega|)$) و درعین‌حال تضمین می‌کند که $\lambda^{[t]}$ در تمامی تکرارها درون سیمپلکس باقی بماند. سوم، ساختار گام فرانک–وولف با فلسفه بهینه‌سازی حداقل–حداکثر حاکم بر چارچوب استوار توزیعی به‌طور طبیعی سازگار است: در هر تکرار، وزن بیشتری به‌صورت تدریجی به تابع زیانی اختصاص می‌یابد که در همان گام، بیشترین مقدار را در میان اعضای مجموعه ابهام Ω به خود اختصاص داده است؛ این رفتار معادل ریاضی همان رویکرد بدبینانه

(کمینه‌سازی بیشینه زیان موردانتظار) است که مسئله (۱۸) بر پایه آن فرمول‌بندی شده است. به‌این‌ترتیب، فرانک–وولف نه‌تنها ازنظر محاسباتی کاراترین انتخاب برای این زیرمسئله خاص است، بلکه تفسیر عملیاتی آن نیز به‌طور مستقیم با معنای نظری بهینه‌سازی استوار توزیعی هم‌راستا است.

بر این اساس، الگوریتم را با $\lambda_\tau^{[0]} = \frac{1}{|\Omega|}$ برای تمام $\tau \in \Omega$ مقداردهی اولیه می‌کنیم. همچنین $p^* = \arg\max_{\tau\in\Omega} ||X - WH||_{2,\tau}$ و $\lambda_*^{[t]}$ را به‌عنوان برداری تعریف می‌کنیم که تنها یک مؤلفه غیرصفر دارد و آن مؤلفه در تکرار t–ام برابر یک بوده و در محل p^* قرار دارد. سپس تمامی λ_τها را در تکرار t+1–ام مطابق رابطه زیر به‌روزرسانی می‌کنیم:

$$\lambda^{[t+1]} = (1-\eta)\lambda^{[t]} + \eta\lambda_*^{[t]} \qquad (۴۲)$$

براساس این فرمول و با توجه به اینکه $||\lambda|| = 1$، مشاهده می‌کنیم که مقدار λ متناظر با بیشینه تابع هزینه در تکرار بعدی افزایش خواهد یافت، درحالی‌که مقادیر مربوط به سایر توابع هزینه کاهش می‌یابند. نرخ این تغییرات توسط پارامتر η کنترل می‌شود؛ به‌طوری‌که مقادیر بزرگ‌تر η باعث تأکید بیشتر بر تابع هزینه بیشینه می‌گردد، و در مقابل، مقادیر کوچک‌تر تغییرات آهسته‌تری در میزان توجه ایجاد می‌کنند. توجه شود که در تمامی حالات، $\eta \in (0,1)$ است. مراحل به‌روزرسانی الگوریتم پیشنهادی DRNMF-bS به شکل شبه‌کد در Algorithm 2 ارائه‌شده است.

۳-۵ تحلیل پیچیدگی محاسباتی

به‌منظور ارزیابی دقیق‌تر هزینه محاسباتی روش پیشنهادی، در این بخش پیچیدگی زمانی مراحل مختلف الگوریتم به‌صورت جزء‌به‌جزء تحلیل و باNMF استاندارد مقایسه می‌شود. همچنین، مرتبه پیچیدگی کلی روش پیشنهادی با روش‌های مقایسه‌ای و پیچیدگی حافظه موردنیاز آن نیز بررسی می‌شود.

الف: تحلیل مرتبه پیچیدگی زمانی

فرض کنید m تعداد ویژگی‌ها، n تعداد نمونه‌ها و r تعداد عوامل نهفته باشد. در مرحله به‌روزرسانی ماتریس وزن قطری D، ابتدا خطای بازسازی هر نمونه محاسبه‌شده و سپس وزن متناظر آن به‌روزرسانی می‌شود. محاسبه خطاها مستلزم تشکیل حاصل‌ضرب WH است که دارای هزینه $O(mnr)$ است. سایر عملیات شامل جمع، تفریق و تقسیم روی عناصر قطری ماتریس تنها هزینه‌ای خطی نسبت به تعداد نمونه‌ها دارند. بنابراین پیچیدگی به‌روزرسانی D از مرتبه $O(mnr)$ خواهد بود.

Algorithm 2. Optimization procedure of the proposed DRNMF-bS algorithm

Input: data matrix X, latent factor r, ambiguity set Ω, regularization parameters α and β;

Output: basis matrix W and representation matrix H;

Initialize W and H randomly, $\lambda_\tau = \frac{1}{|\Omega|}, \forall\tau \in \Omega$;
Compute similarity graph G, $S = G$;
while convergence not reached **do**
for i=1 to n:
Computed $d_i^{(\Omega)}$ according to (26);
end for;
Update instance weight matrix D using $D_{ii} = d_i^{(\Omega)}$;
Update basis matrix W according to (34);
Update representation matrix H according to (35);
Update weights λ according to (42);
Update matrix S using (40) and (41) alternately;
end while
return W and H;

در به‌روزرسانی ماتریس پایه W، عملیات غالب شامل محاسبه عبارت‌های XDH^T و $WHDH^T$ است. ازآنجاکه D یک ماتریس قطری است، ضرب آن تنها به مقیاس‌بندی ستون‌ها محدودشده و هزینه اضافی قابل‌توجهی ایجاد نمی‌کند. درنتیجه پیچیدگی به‌روزرسانی W برابر با $O(mnr)$ است. به‌طور مشابه، در مرحله به‌روزرسانی ماتریس H هم پیچیدگی از مرتبه $O(mnr)$ خواهد بود.

برای به‌روزرسانی ماتریس شباهت دوتصادفی S، ابتدا ماتریس $M = G + \frac{\alpha}{2\beta}H^TH$ محاسبه می‌شود. هزینه تشکیل H^THبرابر با $O(n^2r)$ است. سپس بر اساس روابط (۴۰) و (۴۱)، تنها عملیات جمع، ضرب ماتریس در بردار، نرمال‌سازی و آستانه‌گذاری عنصر به عنصر انجام می‌شود که همگی دارای هزینه $O(n^2)$ هستند. بنابراین پیچیدگی غالب مرحله به‌روزرسانی S برابر با $O(n^2r)$ است. درنتیجه، پیچیدگی محاسباتی هر تکرار الگوریتم پیشنهادی برابر با $O(mnr + n^2r)$ خواهد بود. اگر تعداد تکرارهای الگوریتم برابر t باشد، پیچیدگی کل روش پیشنهادی برابر با $O(t(mnr + n^2r))$ است. جدول ۱ پیچیدگی مراحل به‌روزرسانی الگوریتم پیشنهادی را با الگوریتم NMF استاندارد مقایسه می‌کند.

جدول (۱): پیچیدگی زمانی الگوریتم پیشنهادی و NMF استاندارد به ازای هر تکرار

	Standard NMF	DRNMF-bS
Update W (MUR)	O(mnr)	**O(mnr)**
Update H (MUR)	O(mnr)	**O(mnr)**
Form residuals/norms	O(mn)	**O(mn)**
Compute weight D	-	**O(mnr)**
Update S	-	**O(n^2r)**
Update λ	-	**O(|Ω|)**
Per-iteration total	O(mnr)	**O(mnr + n^2r)**

همچنین برای اینکه کارایی الگوریتم پیشنهادی ازنظر مرتبه زمانی بهتر مشخص شود، مرتبه پیچیدگی زمانی نهایی DRNMF-bS با روش‌های مقایسه‌ای در جدول ۲ گزارش‌شده است.

جدول (۲): مقایسه مرتبه پیچیدگی الگوریتم پیشنهادی با روش‌های مقایسه‌ای

الگوریتم	مرتبه پیچیدگی	الگوریتم	مرتبه پیچیدگی
Frobenius	O(mnr)	Huber	**O(mnr)**
$L_{2,1}$	O(mnr)	Elastic	**O(mnr + n^2r)**
Cauchy	O(mnr^2)	DANMF	**O(lmnr + mn)**
EWNMF	O(mnr)	DRNMF	**O(mn^2r)**
rCIM	O(mnr)	SPLNMF	**O(pmn^2)**
Proposed DRNMF-bS		**O(mnr + n^2r)**	

ب: تحلیل مرتبه پیچیدگی حافظه

به‌منظور ارزیابی هزینه محاسباتی روش پیشنهادی، در این بخش پیچیدگی حافظه موردنیاز مؤلفه‌های اصلی الگوریتم بررسی می‌شود. با توجه به ابعاد ماتریس داده X و ماتریس‌های W و H، حافظه موردنیاز برای ذخیره‌سازی این ماتریس‌ها به‌ترتیب از مرتبه $O(nm)$، $O(rm)$ و $O(rn)$ است. هم‌چنین، ماتریس وزن قطری D و ماتریس درجه U_S در صورت استفاده از نمایش برداری برای عناصر قطری، تنها به حافظه‌ای از مرتبه $O(n)$ نیاز دارند. از سوی دیگر، در محاسبه عبارت منظم‌ساز گراف، با بازنویسی روابط برحسب ماتریس شباهت S و بردار درجه، نیازی به تشکیل صریح ماتریس لاپلاسی L_S وجود ندارد.

در بخش یادگیری گراف، ماتریس شباهت دوتصادفی S و متغیر کمکی M دارای ابعاد $n \times n$ هستند و درنتیجه، در صورت ذخیره‌سازی به‌صورت چگال، هر یک به حافظه‌ای از مرتبه $O(n^2)$ نیاز دارند. با در نظر گرفتن تمامی مؤلفه‌های روش پیشنهادی، پیچیدگی کل حافظه DRNMF-bS را می‌توان به‌صورت $O(nm + r(n+m) + n^2)$ در مقایسه با پیچیدگی حافظه NMF استاندارد با پیچیدگی $O(nm + r(n+m))$ بیان کرد. تفاوت اصلی در مرتبه حافظه دو روش به مؤلفه مربوط به یادگیری و نگهداری ماتریس‌های گراف در DRNMF-bS بازمی‌گردد.

درمجموع، تحلیل فوق نشان می‌دهد که روش پیشنهادی در مقایسه با NMF استاندارد، به‌دلیل به‌روزرسانی وزن‌های استوار و یادگیری ماتریس شباهت دوتصادفی، هزینه محاسباتی و حافظه بیشتری دارد. این هزینه اضافی ناشی از مؤلفه‌های استوار و ساختاری افزوده‌شده به چارچوب NMF است. برای بررسی هزینه محاسباتی از دیدگاه عملی نیز، زمان اجرای DRNMF-bS و تمامی روش‌های مقایسه‌ای بر روی مجموعه‌داده‌های مورد آزمایش اندازه‌گیری و در بخش ٤-٩ گزارش‌شده است. بدین ترتیب، ارزیابی پیچیدگی روش پیشنهادی هم از دیدگاه نظری و هم ازنظر زمان اجرای واقعی انجام‌شده است.

۳-٦ مبانی نظری همگرایی

برای تحلیل دقیق‌تر رفتار الگوریتم، تابع هدف رابطه (۲۹) را به‌صورت زیر نمادگذاری می‌کنیم:

$$\begin{aligned}\varphi(\mathrm{W},\mathrm{H},\mathrm{S},\lambda) = \sum_{\tau\in\Omega}\lambda_\tau\left||X-WH|\right|_{2,\tau} \\ + \alpha Tr(HL_SH^T) \\ + \beta||S-G||_F^2 \\ s.t. W,H,\ \lambda,S \geq 0, \left||\lambda|\right|_1 = 1,\ S \\ = S^T, S1_n = 1_n, diag(S) \\ = 0\end{aligned} \qquad (٤٣)$$

با این نماد، الگوریتم پیشنهادی را می‌توان به‌صورت یک طرح دوگامه در هر تکرار t توصیف کرد:

گام کمینه‌سازی (min-step): با ثابت نگه‌داشتن $\lambda^{[t]}$، سه‌گانه (W,H,S) طی یک دور به‌روزرسانی بلوکی (روابط ٣٤، ٣٥ و ٤٠-٤١) به‌روز می‌شود.

گام بیشینه‌سازی (max-step): با ثابت نگه‌داشتن سه‌گانه $(W^{[t+1]},H^{[t+1]},S^{[t+1]})$، بردار λ طبق رابطه (٤٢) با یک گام فرانک-وولف به‌روز می‌شود.

این ساختار دقیقاً یک طرح کاهش-افزایش تناوبی برای یک مسئله حداقل-حداکثر است که در آن مؤلفه min غیرمحدب و مؤلفه

max، به‌دلیل خطی‌بودن φ نسبت به λ روی سیمپلکس، مقعر (در واقع خطی) است.

گزاره ۱ (کاهش یکنواخت تابع هدف در گام W و H)

فرض: توابع زیان به‌کاررفته در این مقاله عبارت‌اند از $\Xi_1(e)$ =، e، $\Xi_2(e) = e^2$ و $\Xi_{Cau}(e) = ln(e^2 + \gamma^2)$. اگر این توابع را به‌صورت تابعی از $u = e^2$ بنویسیم،یعنی $\varphi_\tau(u) := \Xi_\tau(\sqrt{u})$، هر سه تابع $\varphi_1(u) = \sqrt{u}$ ، $\varphi_2(u) = u$ و $\varphi_{Cau}(u) = \ln(u + \gamma^2)$ روی $u \geq 0$ **مقعر** هستند.

گزاره: برای هر $u \geq 0$ و هر نقطه مرجع $u_0 > 0$، به‌دلیل مقعربودن φ_τ، نامساوی مماس زیر برقرار است:

$$\varphi_\tau(u) \leq \varphi_\tau(u_0) + \psi_\tau(\sqrt{u_0}).(u - u_0) \qquad (٤٤)$$

که در آن ψ_τ همان تابع وزن تعریف‌شده در رابطه (۸) است (و دقیقاً معادل روابط D در (۱٤) و (۱۷) است). درنتیجه، رابطه (٤٤) یک اکثریت‌سازِ نیم‌درجه‌دو (half-quadratic majorizer) برای هر یک از سه تابع زیان فراهم می‌کند [52]. بنابراین، تابع وزن‌دار Q که در رابطه (۲٦) با وزن‌های $d_i^{(\Omega)}$ محاسبه‌شده از تکرار $t-1$ (طبق رابطه ۲۳) تعریف می‌شود، یک اکثریت‌ساز سراسری برای $\varphi_\lambda(W,H,S)$ است، یعنی:

$$\varphi_\lambda(W,H,S) \leq \varphi_\lambda(W,H,S|W^{t-1},H^{t-1}),\forall(W,H,S) \qquad (٤٥)$$

از سوی دیگر، طبق [13]، قواعد به‌روزرسانی ضربی (۳٤) و (۳۵) خود بر پایه یک تابع کمکی برای زیرمسئله مربعی وزن‌دار (۲٦) به‌دست آمده‌اند و تضمین می‌کنند که:

$$Q\left(W^{[t]},H^{[t]}\middle|W^{[t-1]},H^{[t-1]}\right) \leq Q\left(W^{[t-1]},H^{[t-1]}\middle|W^{[t-1]},H^{[t-1]}\right) \qquad (٤٦)$$

با ترکیب (٤۵) و (٤٦) داریم:

$$\begin{aligned}&\varphi_\lambda\left(W^{[t]},H^{[t]},S^{[t-1]}\right)\\ &\leq Q\left(W^{[t]},H^{[t]}\middle|W^{[t-1]},H^{[t-1]}\right)\\ &\leq Q\left(W^{[t-1]},H^{[t-1]}\middle|W^{[t-1]},H^{[t-1]}\right)\\ &= \varphi_\lambda\left(W^{[t-1]},H^{[t-1]},S^{[t-1]}\right)\end{aligned} \qquad (٤٧)$$

یعنی هر گام به‌روزرسانی W یا H، تابع هدف اصلی (غیردرجه‌دوی) φ_λ را کاهش می‌دهد یا ثابت نگه می‌دارد.

گزاره ۲ (زیرمسئله S)

مسئله (۳۷) کمینه‌سازی یک تابع محدب _F^2 ⟦||S-M||⟧ روی اشتراک دو مجموعه محدب است:

الف) زیرفضای وابسته $C_1 = \{S: S = S^T, S1_n = 1_n\}$

ب) مخروط محدب $C_2 = \{S: S \geq 0, diag(S) = 0\}$

طرح تکراری (٤۰)-(٤۱) نمونه‌ای از الگوریتم کلاسیک تصویر متناوب (alternating projection) است که همگرایی دنباله S به یک نقطه شدنی در C_1∩C_2 را تضمین می‌کند [۵۳]. ازآنجاکه این تناوب «تا رسیدن به همگرایی» درون هر تکرار خارجی اجرا می‌شود، S^([t]) را می‌توان کمینه‌کننده زیرمسئله محدب (۳۷) در نظر گرفت؛ درنتیجه، این گام نیز φ_λ را نسبت به بلوک S افزایش نمی‌دهد.

گزاره ۳ (رفتار تابع هدف در گام λ)

با ثابت نگه‌داشتن (W,H,S)، تابع φ نسبت به λ خطی است؛ با تعریف $c_\tau = ||X - WH||_{2,\tau}$، رابطه $\varphi(\lambda) = \sum_\tau \lambda_\tau c_\tau +$ $constant$ برقرار خواهد بود. گام فرانک–وولف در رابطه (٤۲) دقیقاً همان قاعده کلاسیک بیشینه‌سازی خطی روی سیمپلکس است [54]، که درآن $\lambda_*^{[t]}$ رأس بیشینه‌ساز $e\{p*\}$ است. ازآنجاکه $\varphi(\lambda^{[t]})$ یک ترکیب محدب از c_τ است، داریم $\varphi(\lambda^{[t]}) \leq max_\tau c_\tau = \varphi(\lambda_*^{[t]})$ و بنابراین:

$$\begin{aligned}\phi(\lambda^{[t+1]}) &= (1-\eta)\varphi(\lambda^{[t]}) + \eta\varphi(\lambda_*^{[t]})\\ &\geq (1-\eta)\varphi(\lambda^{[t]})\\ &+ \eta\varphi(\lambda^{[t]}) = \varphi(\lambda^{[t]})\end{aligned} \qquad (٤٨)$$

یعنی گام به‌روزرسانی λ، تابع هدف را نسبت به (W, H, S) لزوماً به شکلی یکنواخت کاهش نمی‌دهد.

گزاره‌های ۱ و ۲ نشان می‌دهند که φ در گام min نسبت به λ افزایش نمی‌یابد، درحالی‌که گزاره ۳ نشان می‌دهد φ در گام max نسبت به (W,H,S) لزوماً به شکلی یکنواخت کاهش نمی‌یابد. درنتیجه تا زمانی که λ به تخصیص تقریباً پایدار خود نرسیده است (تکرارهای اولیه الگوریتم)، سهم گام max می‌تواند بر کاهش حاصل از گام min غالب شود و این می‌تواند به شکل افزایش مقدار خطا در گام‌های اولیه خود را نشان دهد. پس‌ازآنکه مدل وزن‌های λ_τ را متناسب با الگوی نوفه داده، به‌اندازه کافی، یاد گرفت آنگاه تکرارهای لزوماً غیرافزایشی مدل شروع می‌شوند. مطالب بیان‌شده و رفتار الگوریتم پیشنهادی به شکل نوسان‌های کوتاه‌مدت اولیه در شکل ۵ نمایش داده‌شده است.

٤- نتایج شبیه‌سازی

در این بخش، به‌منظور ارزیابی جامع میزان مقاومت و کارایی روش پیشنهادی، آزمایش‌های گسترده‌ای بر روی ۱۲ مجموعه‌داده انجام داده و نتایج با ۱۰ روش پایه و پیشرفته مبتنی بر NMF مقایسه می‌شود. برای تضمین مقایسه‌ای منصفانه و قابل‌اعتماد، هر روش با ۱۰ مقداردهی اولیه تصادفی متفاوت اجراشده و نتایج به‌صورت میانگین گزارش‌شده‌اند تا اثر حساسیت به مقداردهی اولیه کاهش یابد. پارامترهای هر روش مطابق با مقادیر پیشنهادی در مقالات اصلی آن تنظیم‌شده‌اند تا سازگاری با مطالعات پیشین حفظ شود. برای ارزیابی خوشه‌بندی، الگوریتم استاندارد K–میانگین بر روی ماتریس H اعمال‌شده و عملکرد با استفاده از دو معیار متداول NMI و ACC سنجیده شده است. این معیارها میزان انطباق برچسب‌های پیش‌بینی‌شده با برچسب‌های واقعی را ارزیابی می‌کنند و تصویری جامع از کیفیت خوشه‌بندی ارائه می‌دهند. بدین ترتیب، مقایسه‌ای مستحکم و معنادار میان روش پیشنهادی و رویکردهای موجود تضمین می‌شود. ماتریس‌ها به‌صورت تصادفی مقداردهی اولیه شده و با قواعد به‌روزرسانی ضربی طی ۳۰۰ تکرار به‌روزرسانی شده‌اند.

٤-١ مجموعه داده‌های مورد آزمایش

روش پیشنهادی بر روی ۱۲ مجموعه‌داده معیار از حوزه‌های تشخیص چهره، تشخیص اشیاء، طبقه‌بندی ارقام دست‌نویس و پوشاک، تحلیل زیستی و تصویربرداری پزشکی ارزیابی‌شده است. مجموعه‌داده‌های چهره شامل تصاویر خاکستری با شرایط نوری و زوایای دید متفاوت هستند که برای افزایش کارایی محاسباتی تغییر اندازه داده‌شده‌اند. مجموعه‌داده‌های MNIST، USPS و Fashion-MNIST برای انجام آزمایش‌های کارآمدتر –با استفاده از نمونه‌برداری تصادفی در هر کلاس– به زیرمجموعه‌هایی با ۱۰۰۰ تا ۱۱۰۰ نمونه کاهش داده‌شده‌اند. مجموعه‌داده‌های زیستی شامل Seeds و Ecoli وظایف طبقه‌بندی متنوعی را پوشش می‌دهند. مجموعه‌داده‌های پزشکی استخراج‌شده از MedMNIST شامل OrganA، Blood و Pneumonia هستند که قابلیت مقیاس‌پذیری و انطباق روش پیشنهادی با داده‌های پربعد و پیچیده را نشان می‌دهند. جزئیات مجموعه‌داده‌ها در جدول ۳ ارائه‌شده است. استفاده از این مجموعه‌داده‌های متنوع، ارزیابی دقیق و جامعی از پایداری و کارایی روش پیشنهادی در کاربردهای واقعی فراهم می‌کند.

جدول (۳): توصیف و ابعاد مجموعه داده‌های مورد آزمایش

مجموعه داده	کاربرد	تعداد نمونه	تعداد دسته	تعداد ویژگی
Yale	تشخیص چهره	۱٦٥	۱٥	۱۰۲٤
ORL	تشخیص چهره	٤۰۰	٤۰	۱۰۲٤
COIL20	تشخیص اشیاء	۱٤٤۰	۲۰	۱۰۲٤
MNIST	ارقام دست‌نویس	۱۰۰۰	۱۰	۷۸٤
Fashion	لباس	۱۰۰۰	۱۰	۷۸٤
Seeds	بیولوژی	۲۱۰	۳	۷
Ecoli	بیولوژی	۳۳٦	۸	۷
USPS	ارقام دست‌نویس	۱۱۰۰	۱۰	۲٥٦
UMIST	تشخیص چهره	٥۷٥	۲۰	٦٤٤
OrganA	سی‌تی‌اسکن شکم	٥۸۸۳۰	۱۱	۷۸٤
Blood	گلبول قرمز	۱۷۰۹۲	۸	۲۳٥۲
Pneumonia	تصویر اشعه ایکس قفسه سینه	٥۸٥٦	۲	۷۸٤

٤-۲ روش‌های مقایسه‌ای

به‌منظور بررسی و تأیید عملکرد روش پیشنهادی DRNMF-bS در نمایش داده‌ها، این روش با ۱۰ مدل مختلف NMF مقایسه شده است که شامل رویکردهای مرسوم، مبتنی بر عنصر، مبتنی بر نمونه و مبتنی بر یادگیری عمیق می‌باشند. این روش‌ها مجموعه‌ای متنوع از تکنیک‌ها را تشکیل می‌دهند که هر یک به شیوه‌ای متفاوت باهدف افزایش پایداری و مقاومت NMF طراحی‌شده‌اند:

- Frobenius-NMF: فرم استاندارد NMF که تابع زیان خطای مربعی را کمینه می‌کند و به‌عنوان مبنای مقایسه برای ارزیابی بهبودهای حاصل از روش‌های استوار به‌کار می‌رود [۱۳].
- L2,1-NMF: یک نسخه استوار از NMF که در آن تابع زیان کمترین مربعات با نرم L2,1 جایگزین شده است [۳۲].
- Cauchy-NMF: یک روش عنصرمحور که با فرض توزیع کوشی برای خطای بازسازی، در برابر نوفه‌های شدید عملکرد مؤثری دارد [۳٤].

- EWNMF: روش NMF با وزن‌دهی آنتروپی که با تخصیص وزن‌های قابل‌بهینه‌سازی، مقاومت در سطح ویژگی را افزایش می‌دهد [۴۷].
- rCIM-NMF: یک روش استوار نمونه‌محور که با بهره‌گیری از معیار کورنتروپی، تأثیر نمونه‌های پرت را در تجزیه کاهش می‌دهد [۴۸].
- Huber-NMF: رویکردی عنصرمحور که با استفاده از تابع هوبر، تعادلی میان خطای مربعی و قدرمطلق ایجاد می‌کند [۴۸].
- Elastic-NMF: مدلی که یک تابع زیان کشسان ارائه می‌دهد و به‌صورت پیوسته بین زیان فروبنیوس و نرم L2,1 گذار می‌کند [۴۹].
- DANMF: یک مدل NMF مبتنی بر یادگیری عمیق که با استفاده از ساختار چندلایه، نمایش‌های سلسله‌مراتبی را استخراج کرده و درعین‌حال خاصیت تفسیرپذیری تجزیه نامنفی را حفظ می‌کند [۵۰].
- DRNMF: یک مدل NMF استوار توزیعی در سطح عنصر که مجموعه‌ای از توزیع‌های نوفه شامل گاوسی، پواسون و گاما را پوشش می‌دهد [۳۹].
- SPLNMF: یک روش استوار NMF که یادگیری خودگام را با L2,1-NMF ترکیب کرده و با اجتناب از کمینه‌های محلی نامطلوب، مقاومت را بهبود می‌بخشد [۵۱].

۴-۳ نتایج خوشه‌بندی

به‌منظور ارزیابی جامع عملکرد روش پیشنهادی، آزمایش‌ها بر روی دوازده مجموعه‌داده پرکاربرد که حوزه‌های کاربردی متنوعی را پوشش می‌دهند انجام‌شده است. این مجموعه‌داده‌ها ازنظر تعداد نمونه‌ها، ابعاد ویژگی‌ها و ساختار خوشه‌ای دارای تنوع قابل‌توجهی هستند و بدین‌ترتیب بستر مناسبی برای سنجش کارایی و قابلیت تعمیم روش پیشنهادی فراهم می‌کنند. مدل پیشنهادی با ۱۰ روش شاخص خوشه‌بندی که نمایانگر رویکردهای مدل‌سازی متنوع و به‌عنوان خطوط مبنای قدرتمند شناخته می‌شوند، مقایسه شده است. عملکرد خوشه‌بندی با استفاده از دو معیار استاندارد اطلاعات متقابل نرمال‌شده (NMI) و دقت خوشه‌بندی (ACC) ارزیابی‌شده است. مقادیر هر دو معیار در بازه [0,1] قرار داشته و مقادیر بزرگ‌تر بیانگر عملکرد بهتر هستند. به‌منظور کاهش تأثیر مقداردهی اولیه تصادفی، هر آزمایش ده بار تکرار شده و میانگین نتایج گزارش‌شده است.

الف: داده‌های بدون نوفه

جدول‌های ٤ و ٥ نتایج عملکرد خوشه‌بندی تمامی روش‌ها را بر روی داده‌های بدون نوفه برحسب معیارهای NMI و ACC گزارش می‌کنند. همان‌گونه که از جدول ٤ مشاهده می‌شود مدل پیشنهادی در اغلب مجموعه‌داده‌ها بهترین عملکرد را به خود اختصاص داده است.

جدول (٤): مقایسه روش پیشنهادی روی مجموعه‌داده‌های بدون نوفه بر اساس معیار NMI

	Frobenius	$L_{2,1}$	Cauchy	EWNMF	rCIM	Huber	Elastic	DANMF	DRNMF	SPLNMF	DRNMF-bS
Yale	0.4597	0.4832	0.4968	0.4712	0.4607	0.4828	0.4756	0.4598	0.5007	0.4802	**0.5365**
ORL	0.7975	0.7847	0.7762	0.7964	0.7663	0.7835	0.7956	0.7326	0.7125	0.7413	**0.8256**
COIL20	0.7575	0.7655	0.7325	0.7988	0.7538	0.7970	0.7021	0.7345	0.6812	0.7661	**0.8021**
MNIST	0.4683	0.4631	0.4559	**0.5080**	0.4900	0.4470	0.4369	0.4108	0.4896	0.4560	0.4978
Fashion	0.5230	0.5353	0.5440	0.5297	0.5353	0.5597	0.5155	0.5654	0.5312	0.5024	**0.6103**
Seeds	0.6321	0.5735	0.5230	0.5023	0.5595	0.6124	0.4730	0.6501	0.5863	0.5587	**0.7145**
Ecoli	0.5358	0.5298	0.5758	0.5024	0.5242	0.5534	0.5008	0.4987	0.5188	0.5951	**0.6320**
USPS	0.3894	0.3072	0.3811	0.3687	0.3958	0.3756	0.3574	0.4088	0.3863	0.3663	**0.4586**
UMIST	0.6023	0.5975	0.5823	0.5987	0.6080	0.5847	0.6259	0.5956	0.5549	0.5873	**0.6423**
OrganA	0.6044	0.6120	0.6272	0.6568	0.6287	0.6041	0.5800	0.6485	0.6124	0.6134	**0.6911**
Blood	0.2963	0.3609	0.2822	0.2496	0.2647	0.2402	0.2689	0.3190	0.2996	0.3050	**0.3987**
Pneumonia	0.3078	0.3078	0.3333	0.3025	0.2541	0.3079	0.3199	0.3003	0.2969	0.2889	**0.3652**

جدول (۵): مقایسه روش پیشنهادی روی مجموعه‌داده‌های بدون نوفه بر اساس معیار ACC

	Frobenius	L2,1	Cauchy	EWNMF	rCIM	Huber	Elastic	DANMF	DRNMF	SPLNMF	DRNMF-bS
Yale	0.4388	0.4445	0.4377	0.4366	0.4494	0.4323	0.4573	0.4920	0.4321	0.4767	**0.5630**
ORL	0.6570	0.6858	0.6725	0.6950	0.6592	0.6040	0.6711	0.5825	0.6002	0.6573	**0.7183**
COIL20	0.6662	0.6586	0.6950	06679	0.6085	0.6882	0.6697	0.6521	0.6332	0.6674	**0.7586**
MNIST	0.5025	0.5032	0.5270	0.5045	0.4355	0.4857	0.5105	0.4850	0.4875	0.5467	**0.5948**
Fashion	0.5785	0.6040	0.5845	0.6145	0.5770	0.5870	0.5600	0.5892	0.5742	0.5123	**0.6753**
Seeds	0.8641	0.8329	0.8026	0.8218	0.8076	0.7962	0.8132	0.8224	0.7863	0.7981	**0.9234**
Ecoli	0.7516	0.7691	0.7515	0.7729	0.7684	0.7189	0.7960	0.7525	0.7745	0.7330	**0.8586**
USPS	0.4619	0.4968	0.4759	0.4468	0.4355	0.4784	0.4082	0.4444	0.4632	0.4636	**0.5546**
UMIST	0.4819	0.4793	0.4783	0.4600	0.4967	0.4708	0.4852	**0.5313**	0.4669	0.4978	**0.5274**
OrganA	0.6533	0.6878	0.6757	0.6745	0.6706	0.6657	0.6820	0.7377	0.6623	0.6728	**0.7684**
Blood	0.4902	0.4884	0.5136	0.4131	0.4036	0.4416	0.4179	0.4329	0.3812	0.4983	**0.5340**
Pneumonia	0.7825	0.7266	0.8062	0.7895	0.7654	0.7712	0.7804	0.7663	0.7348	0.7744	**0.8464**

این نتایج بیانگر توانایی بالای مدل در بهره‌برداری مؤثر از اطلاعات موجود و یادگیری یک بازنمایی مؤثر است. برتری DRNMF-bS به‌ویژه درمجموعه‌داده‌هایی با ساختار پیچیده‌تر مشهودتر است که نشان می‌دهد ترکیب هم‌زمان یادگیری استوار توزیعی و لحاظ‌کردن منظم‌ساز دوتصادفی می‌تواند به تفکیک خوشه‌ای دقیق‌تر منجر شود. روند مشابهی در معیار ACC نیز مشاهده می‌شود؛ به‌طوری‌که DRNMF-bS در اکثر مجموعه‌داده‌ها بالاترین مقادیر را کسب کرده است. این موضوع نشان می‌دهد که انتساب‌های خوشه‌ای حاصل از این مدل از انسجام درونی بیشتر و پایداری بالاتری برخوردارند.

ب: داده‌های نوفه‌دار

به‌منظور ارزیابی میزان پایداری و استحکام روش پیشنهادی در شرایط واقع‌بینانه و چالش‌برانگیز، سناریوی خوشه‌بندی در حضور نوفه نیز موردبررسی قرارگرفته است. در این تنظیم آزمایشی، نوفه اضافه‌شده از میان توزیع‌های گاوسی، لاپلاس و کوشی به‌صورت تصادفی نمونه‌برداری شده است. درنتیجه، نمونه‌های هر مجموعه‌داده می‌تواند به انواع متفاوتی از نوفه آلوده شوند و توزیع نوفه بین مجموعه‌داده‌ها نه یکسان است و نه از پیش معلوم. این پروتکل آزمایشی بازتاب‌دهنده تعریف واقعی یادگیری استوار توزیعی است؛ چراکه در عمل، توزیع دقیق نوفه معمولاً ناشناخته بوده و تنها می‌توان فرض کرد که درمجموعه‌ای از توزیع‌های محتمل قرار دارد.

در چنین شرایطی، عدم قطعیت در داده‌ها تشدید می‌شود، زیرا هیچ تضمینی وجود ندارد که تمامی مجموعه‌داده‌ها تحت یک سازوکار آلودگی مشترک قرارگرفته باشند. بسیاری از روش‌های موجود خوشه‌بندی مبتنی بر فرض نوفه همگن یا استفاده از تابع زیان ثابت هستند و درنتیجه توانایی محدودی در مواجهه با نوفه‌های ناهمگن و ناشناخته دارند. در مقابل، چارچوب پیشنهادی با بهره‌گیری از صورت‌بندی استوار توزیعی و وزن‌دهی تطبیقی زیان، به‌طور صریح برای مقابله با این ناهمگنی طراحی‌شده است. این سازوکار امکان سازگاری خودکار مدل با ویژگی‌های آماری متفاوت نوفه‌های مختلف را بدون نیاز به اطلاع قبلی از نوع توزیع نوفه فراهم می‌سازد. همان‌گونه که نتایج جداول ٦ و ٧ در شرایط نوفه‌ای نشان می‌دهد، مدل پیشنهادی نسبت به سایر روش‌های مقایسه‌ای افت عملکرد کمتری داشته و پایداری بالاتری را حفظ می‌کند که این امر بیانگر نوآوری و مزیت اصلی رویکرد پیشنهادی در مدیریت ناهمگنی توزیعی است.

جدول (٦): مقایسه روش پیشنهادی روی مجموعه‌داده‌های نوفه‌ای بر اساس معیار NMI

	Frobenius	$L_{2,1}$	Cauchy	EWNMF	rCIM	Huber	Elastic	DANMF	DRNMF	SPLNMF	DRNMF-bS
Yale	0.4137	0.4201	0.4099	0.4446	0.4153	0.4280	0.4098	0.4299	0.4002	0.4598	**0.5245**
ORL	0.7314	0.7287	0.7125	0.7601	0.7240	0.7251	0.6976	0.7066	0.7136	0.7556	**0.8197**
COIL20	0.6922	0.7080	0.7133	0.7369	0.6833	0.6707	0.6855	0.7275	0.6995	0.7124	**0.7978**
MNIST	0.4234	0.4004	0.4137	0.3933	0.4010	0.3979	0.4318	0.4067	0.3846	0.4103	**0.4754**
Fashion	0.5175	0.5494	0.5234	0.4976	0.5014	0.5137	0.5256	0.5393	0.4996	0.5182	**0.6049**
Seeds	0.4453	0.4394	0.5020	0.5201	0.4640	0.5372	0.4516	0.5700	0.5321	0.5674	**0.6943**
Ecoli	0.5160	0.4993	0.5132	0.5174	0.5109	0.5057	0.5170	0.5316	0.5112	0.5068	**0.6097**
USPS	0.3715	0.3615	0.3519	0.3614	0.3134	0.3209	0.3047	0.3444	0.3612	0.3149	**0.4401**
UMIST	0.5941	0.5917	0.6064	0.5587	0.5469	0.5913	0.5328	0.5786	0.5547	0.5494	**0.6336**
OrganA	0.5607	0.5640	0.5854	0.5631	0.5009	0.5741	0.5651	0.5090	0.5447	0.6181	**0.6870**
Blood	0.2460	0.2573	0.2968	0.2873	0.2831	0.2726	0.2851	0.3043	0.2875	0.2926	**0.3843**
Pneumonia	0.2455	0.1981	0.1832	0.2752	0.2519	0.1944	0.2691	0.2374	0.2014	0.2357	**0.3657**

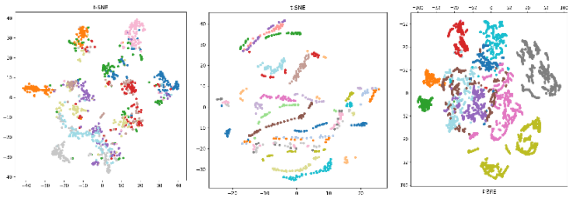
جدول (٧): مقایسه روش پیشنهادی روی مجموعه‌داده‌های نوفه‌ای بر اساس معیار ACC

	Frobenius	$L_{2,1}$	Cauchy	EWNMF	rCIM	Huber	Elastic	DANMF	DRNMF	SPLNMF	DRNMF-bS
Yale	0.3980	0.4141	0.4424	0.4602	0.4182	0.4182	0.3942	0.4162	0.3546	0.3984	**0.5548**
ORL	0.6354	0.6463	0.6488	0.6575	0.6463	0.6687	0.6732	0.4975	0.6177	0.6463	**0.7000**
COIL20	0.6701	0.6587	0.6812	0.6438	0.6306	0.6556	0.6576	0.6756	0.5847	0.6256	**0.7474**
MNIST	0.4980	0.5207	0.4920	0.5060	0.4953	0.4897	0.5037	0.5350	0.5139	0.5693	**0.5820**
Fashion	0.5560	0.5994	0.5880	0.5650	0.5609	0.4975	0.5635	0.5020	0.5054	0.5675	**0.6700**
Seeds	0.8078	0.7462	0.7814	0.7971	0.6914	0.8095	0.7762	0.8400	0.7659	0.8637	**0.9062**
Ecoli	0.6969	0.7285	0.7150	0.7542	0.7646	0.7455	0.7494	0.7408	0.7321	0.7520	**0.8157**
USPS	0.4236	0.4230	0.4627	0.4776	0.4279	0.4297	0.4052	0.4590	0.3902	0.4298	**0.5499**
UMIST	0.4367	0.4704	0.5096	0.4765	0.4626	0.4157	0.4835	0.3178	0.4775	0.4991	**0.5196**
OrganA	0.5897	0.5757	0.6360	0.5863	0.5467	0.5391	0.6268	0.5655	0.4865	0.5438	**0.7175**
Blood	0.4151	0.4162	0.4363	0.4293	0.4211	0.3957	0.4033	0.4698	0.4126	0.4579	**0.5241**
Pneumonia	0.7671	0.7701	0.7243	0.7815	0.7987	0.6913	0.7615	0.7423	0.6895	0.7143	**0.8372**

٤-٤ آزمون t-SNE

در شکل ۱، تجسم دوبعدی ویژگی‌های نهفته (H) استخراج‌شده توسط مدل پیشنهادی با استفاده از روش t-SNE برای مجموعه‌داده‌های OrganA، UMIST و MNIST ارائه‌شده است. همان‌گونه که مشاهده می‌شود، نمونه‌ها در فضای نهفته به‌صورت خوشه‌هایی نسبتاً منسجم و قابل تفکیک سازمان‌دهی شده‌اند که نشان‌دهنده توانایی مدل در یادگیری بازنمایی‌های معنادار برای اهداف خوشه‌بندی است. اگرچه مرز میان برخی خوشه‌ها کاملاً شفاف و ایده‌آل نیست، اما ساختار کلی داده‌ها به‌خوبی حفظ شده و انسجام درون‌خوشه‌ای در سطح قابل‌قبولی قابل مشاهده است. عدم جداسازی کامل خوشه‌ها را می‌توان به ماهیت خطی NMF و ظرفیت محدود آن در مدل‌سازی ساختارهای غیرخطی و خمینه داده‌ها نسبت داد. با توجه به اینکه تمرکز اصلی این پژوهش بر طراحی و تحلیل تابع زیان استوار توزیعی بوده است، توسعه بخش بازنمایی و افزایش پیچیدگی مدل در این کار مدنظر قرار نگرفته است. در پژوهش‌های آتی می‌توان با بهره‌گیری از نسخه‌های

منظم‌سازی‌شده مبتنی بر گراف یا خمینه، ساختار هندسی داده‌ها را بهتر لحاظ کرده و به جداسازی واضح‌تر خوشه‌ها در فضای نهفته دست‌یافت.



الف ب ج

شکل (۱): نمودار tSNE الگوریتم پیشنهادی

الف) OrganA ب) UMIST ج) MNIST

۵-٤ مقایسه iDRNMF با DRNMF-bS

به‌منظور تبیین جایگاه و نوآوری مدل پیشنهادی، در این بخش عملکرد روش پایه iDRNMF با نسخه توسعه‌یافته آن، یعنی DRNMF-bS، به‌صورت جامع مقایسه می‌شود. آزمایش‌ها بر روی ۱۲ مجموعه‌داده انجام‌شده و نتایج با استفاده از دو معیار استاندارد خوشه‌بندی، یعنی ACC و NMI، ارزیابی‌شده‌اند. مقادیر مربوطه در دو زیرشکل از شکل ۲ ارائه‌شده‌اند.

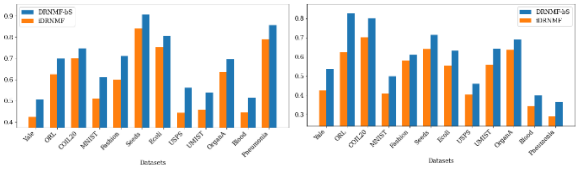
مدل iDRNMF با تکیه بر چارچوب استوار توزیعی در سطح نمونه، قادر است اثر نوفه‌های ناهمگن و عدم‌قطعیت‌های توزیعی را تا حد قابل‌توجهی کاهش دهد. این مدل از طریق وزن‌دهی تطبیقی زیان‌ها، نمونه‌های ناسازگار با توزیع غالب داده را تعدیل می‌کند و درنتیجه، در برابر انواع نوفه عملکردی پایدار ارائه می‌دهد. بااین‌حال، تمرکز اصلی iDRNMF بر جنبه آماریِ مقاومت در برابر نوفه است و سازوکار مستقلی برای اعمال قیود ساختاری بر هندسه فضای نهفته در اختیار ندارد. درنتیجه، اگرچه بازسازی داده‌ها به‌شکل استوار انجام می‌شود، اما ساختار همسایگی و انسجام سراسری نمایش‌های نهفته به‌صورت صریح کنترل نمی‌گردد.

در مقابل، DRNMF-bS با افزودن تنظیم‌گر دو-تصادفی به چارچوب استوار توزیعی، یک قید ساختاری قدرتمند بر ماتریس نمایش نهفته تحمیل می‌کند. این تنظیم‌گر با واداشتن ساختار ارتباطی به رعایت خاصیت دو-تصادفی، توزیعی متعادل از وابستگی‌ها میان نمونه‌ها ایجاد کرده و از تمرکز نامتوازن وزن‌ها بر زیرمجموعه‌ای محدود جلوگیری می‌کند. درنتیجه، فضای نهفته حاصل نه‌تنها ازنظر آماری استوار است، بلکه از دیدگاه هندسی نیز منسجم، پایدار و تفکیک‌پذیرتر می‌شود.

مطابق نتایج شکل ۲، مدل DRNMF-bS در همه مجموعه‌داده‌ها و در هر دو معیار ACC و NMI بهبود معناداری نسبت به iDRNMF نشان می‌دهد. این بهبود هم‌زمان در دو شاخص مکمل، بیانگر آن است که ارتقاء حاصل صرفاً یک افزایش عددی محدود در دقت برچسبی نیست، بلکه بیانگر بهبود ساختار اطلاعاتی و کیفیت کلی خوشه‌بندی است. به‌عبارت دیگر، نمایش‌های تولیدشده توسط DRNMF-bS هم ازنظر تطابق با برچسب‌های واقعی و هم ازنظر حفظ ساختار خوشه‌ای، کیفیت بالاتری دارند.

از منظر مفهومی، این نتایج نشان می‌دهد که افزودن تنظیم‌گر دو-تصادفی صرفاً یک جمله منظم‌ساز اضافی نیست، بلکه یک لایه ساختاری مکمل برای چارچوب استوار توزیعی ایجاد می‌کند. در حالی‌که iDRNMF بر «چه میزان اعتماد به هر نمونه» تمرکز دارد، DRNMF-bS علاوه بر آن، «چگونگی توزیع ساختار ارتباطی در فضای نهفته» را نیز کنترل می‌کند. این تمایز، ارتقاء مدل را از سطح یک بهبود پارامتری به سطح یک تقویت ساختاری و مفهومی ارتقاء می‌دهد.

پایداری الگوی برتری DRNMF-bS در تمامی ۱۲ مجموعه‌داده نیز حاکی از آن است که اثر تنظیم‌گر دو-تصادفی وابسته به ویژگی خاص یک داده مشخص نیست، بلکه یک مکانیزم عمومی برای تقویت انسجام نمایش نهفته فراهم می‌آورد. بدین ترتیب، DRNMF-bS ضمن حفظ مزایای استوار توزیعی iDRNMF، محدودیت ساختاری آن را برطرف کرده و به نمایش‌هایی متوازن‌تر، استوارتر و با قدرت تفکیک بالاتر دست می‌یابد.



شکل (۲): تأثیر مؤلفه تنظیم‌گر دوتصادفی

مقایسه مستقیم iDRNMF و DRNMF-bS نشان می‌دهد که تفاوت دو روش صرفاً در اضافه شدن یک جمله منظم‌ساز به تابع هدف خلاصه نمی‌شود. iDRNMF مسئله عدم‌قطعیت آماری را با

یادگیری تطبیقی وزن توابع زیان حل می‌کند، اما ساختار روابط نمونه‌ها را ثابت و خارج از فرآیند اصلی یادگیری در نظر می‌گیرد. در DRNMF-bS، این محدودیت با یادگیری هم‌زمان ماتریس شباهت دوتصادفی برطرف می‌شود. بنابراین، bS-DRNMF قابلیت مقاومت آماری موجود در iDRNMF را حفظ کرده و یک لایه ساختاری جدید به آن اضافه می‌کند. به بیان ساده، iDRNMF به این پرسش پاسخ می‌دهد که «در برابر الگوهای مختلف خطای بازسازی، چه میزان وزن به هر تابع زیان داده شود؟»، درحالی‌که DRNMF-bS علاوه بر آن پاسخ می‌دهد که «روابط میان نمونه‌ها در فضای نهفته چگونه باید اصلاح و حفظ شوند؟»

٤-٦ مطالعه حذف مولفه‌ها

به‌منظور بررسی نقش مؤلفه‌های مختلف روش پیشنهادی و ارزیابی سهم هر یک از آن‌ها در عملکرد نهایی، مطالعه حذف مؤلفه‌ها در دو سطح انجام‌شده است. در بخش نخست، اثر مؤلفه‌های اصلی مدل پیشنهادی به‌صورت جداگانه بررسی می‌شود و در بخش دوم، تأثیر انتخاب مجموعه ابهام در جمله استوار توزیعی مورد ارزیابی قرار می‌گیرد.

الف: بررسی اثر مؤلفه‌های اصلی مدل پیشنهادی

به‌منظور بررسی نقش و سهم هر یک از مؤلفه‌های کلیدی در مدل پیشنهادی DRNMF-bS، مطالعه حذف مؤلفه (Ablation Study) جامع تحت چهار تنظیم مختلف انجام‌شده است. عملکرد خوشه‌بندی بر اساس دو معیار ACC و NMI بر روی پنج مجموعه‌داده معیار شامل ORL، MNIST، Seeds، OrganA و Blood گزارش‌شده که نتایج آن در شکل ۳ قابل مشاهده است. هر ستون در این شکل متناظر با یکی از پیکربندی‌های زیر است:

تنظیم ۱: در این حالت، تمامی جمله‌های منظم‌ساز حذف شده و مدل به فرم پایه استوار توزیعی تقلیل می‌یابد.

تنظیم ۲: در این تنظیم، تنظیم گر دو تصادفی بدون اعمال قید وابستگی به ساختار گراف اولیه یاد گرفته می‌شود.

تنظیم ۳: ضرایب وزن‌دهی زیان‌ها (λ) ثابت در نظر گرفته می‌شوند و سازوکار تطبیقی انتخاب زیان استوار توزیعی غیرفعال می‌گردد؛ درنتیجه مدل به یک چارچوب چندهدفه ایستا تبدیل می‌شود.

تنظیم ٤: نسخه کامل پیشنهادی شامل وزن‌دهی تطبیقی زیان‌های استوار توزیعی به‌همراه منظم‌سازی دو تصادفی.

شایان ذکر است حالتی که در آن $\alpha = 0$ ازآنجاکه عملاً تأثیری بر ماتریس H ندارد، گزارش نشده است.

این مطالعه باهدف بررسی این موضوع طراحی‌شده است که آیا بهبود عملکرد صرفاً ناشی از استفاده هم‌زمان از چند تابع زیان است یا اینکه سازوکار یادگیری تطبیقی وزن‌ها و یادگیری ساختار گراف، هر یک به‌صورت مستقل و مکمل در عملکرد مدل نقش دارند. به‌طور خاص، در تنظیم سوم، وزن‌های λ ثابت نگه‌داشته می‌شوند تا اثر «ترکیب چند تابع زیان با ضرایب ثابت» از اثر «وزن‌دهی تطبیقی استوار توزیعی» تفکیک شود. بنابراین، این آزمایش مستقیماً نشان می‌دهد که مزیت مدل پیشنهادی را نمی‌توان صرفاً به وجود چند تابع زیان نسبت داد.

همان‌گونه که در شکل ۳ مشاهده می‌شود، مدل کامل DRNMF-bS(تنظیم ٤) به‌طور پایدار بهترین عملکرد را در تمامی مجموعه‌داده‌ها و در هر دو معیار ACC و NMI کسب کرده است. این نتیجه نشان می‌دهد که تلفیق هم‌زمان وزن‌دهی تطبیقی استوار توزیعی و منظم‌سازی دوتصادفی نقش تعیین‌کننده‌ای در ارتقای کیفیت خوشه‌بندی دارد. بهبودهای حاصل در میان مجموعه‌داده‌های مختلف پایدار هستند که بیانگر آن است این دو مؤلفه اثر مکمل داشته و عملکردی هم‌افزا ایجاد می‌کنند.

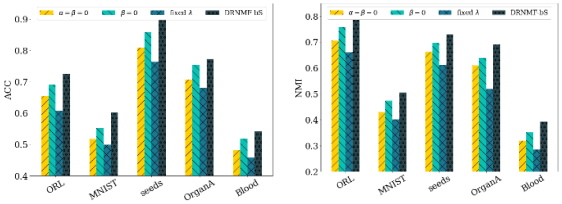
رتبه دوم عموماً به تنظیم ۲ اختصاص دارد؛ هرچند این نسخه همچنان نسبت به مدل پایه بهبود نشان می‌دهد، افت عملکرد آن نسبت به مدل کامل بیانگر اهمیت اعمال قیود دوتصادفی در ایجاد ساختاری پایدار، جلوگیری از تخصیص‌های نوفه‌ای، و حفظ سازگاری سراسری نمایش نهفته است. این موضوع نشان می‌دهد که منظم‌سازی دوتصادفی نقش مهمی در متعادل‌سازی انعطاف‌پذیری مدل و انسجام ساختاری ایفا می‌کند.

تنظیم ۱ در رتبه سوم قرار می‌گیرد. این نتیجه تأیید می‌کند که خودِ چارچوب استوار توزیعی نیز به‌تنهایی قادر به بهبود عملکرد از طریق مقابله با نوفه‌های ناهمگن و عدم‌قطعیت‌های توزیعی است. بااین‌حال، در غیاب منظم‌سازی ساختاری، نمایش‌های آموخته‌شده

از قدرت تفکیک‌پذیری کمتری برخوردارند، به‌ویژه درمجموعه‌داده‌هایی با ساختار پیچیده یا پراکندگی بالا.

درنهایت، تنظیم ۳ ضعیف‌ترین عملکرد را در هر دو معیار ACC و NMI نشان می‌دهد. باوجود حفظ چندین جمله زیان، ثابت نگه‌داشتن ضرایب λ سازوکار اصلی یادگیری استوار توزیعی را غیرفعال می‌کند و مانع از انطباق پویا با الگوهای نوفه‌ای متفاوت می‌شود. درنتیجه، مدل به یک مسأله بهینه‌سازی چندهدفه ایستا تقلیل می‌یابد که توانایی کافی برای مقابله با عدم‌قطعیت‌های ناشناخته را ندارد. جالب‌توجه آن‌که عملکرد این حالت حتی از مدل پایه نیز ضعیف‌تر است که نشان می‌دهد مدل‌سازی نادرست یا غیرتطبیقیِ استوار توزیعی می‌تواند اثر معکوس داشته باشد.

این نتیجه ازنظر تبیین نوآوری روش اهمیت ویژه‌ای دارد. اگر عملکرد بهتر مدل تنها ناشی از ترکیب چند تابع زیان بود، انتظار می‌رفت نسخه دارای ضرایب ثابت نیز عملکردی نزدیک به مدل کامل داشته باشد. بااین‌حال، نتایج نشان می‌دهند که ثابت نگه‌داشتن λ باعث افت عملکرد می‌شود. بنابراین، عامل مؤثر تنها «تعداد توابع زیان» نیست، بلکه یادگیری تطبیقی اهمیت آن‌ها در چارچوب استوار توزیعی نقش اساسی دارد. درعین‌حال، مقایسه تنظیم‌های مختلف نشان می‌دهد که افزودن گراف دوتصادفی نیز یک مؤلفه مستقل و مکمل است. درنتیجه، نتایج تجربی از این دیدگاه پشتیبانی می‌کنند که مدل پیشنهادی حاصل یک ترکیب ساده و ایستا از اجزای موجود نیست، بلکه از تعامل میان جمله استوار توزیعی تطبیقی و یادگیری ساختار گراف بهره می‌برد.



شکل (۳): تحلیل حذف مولفه‌ها در الگوریتم پیشنهادی

ب: بررسی اثر مجموعه ابهام در جمله استوار توزیعی

به‌منظور تکمیل مطالعه حذف مؤلفه‌ها و بررسی دقیق‌تر سهم جمله استوار توزیعی در عملکرد مدل پیشنهادی **DRNMF-bS**، یک آزمایش حذف مؤلفه تکمیلی با تمرکز بر ساختار مجموعه ابهام انجام‌شده است. در این آزمایش، به‌جای درنظرگرفتن همزمان سه تابع زیان L_1 فروبنیوس و کوشی، مجموعه ابهام به‌صورت تدریجی گسترش داده شد تا مشخص شود افزودن هر تابع زیان چه تأثیری بر توانایی مدل در مواجهه با انواع مختلف نوفه دارد. لازم به ذکر است که چارچوب استوار توزیعی پیشنهادی به یک مجموعه ابهام خاص محدود نیست و می‌تواند متناسب با دانش پیشین یا ویژگی‌های مسئله، مجموعه‌ای از توابع زیان با ویژگی‌های آماری متفاوت را دربرگیرد. در این مطالعه، مجموعه ابهام متشکل از L_1 فروبنیوس و کوشی به‌عنوان یک نمونه نماینده در نظر گرفته‌شده است؛ به‌گونه‌ای که این سه تابع زیان به‌ترتیب ویژگی‌های متفاوتی از نوفه‌های تنک/لاپلاسی، گاوسی و دارای دنباله‌های سنگین را پوشش می‌دهند.

بر این اساس، هفت پیکربندی مختلف از مجموعه ابهام بررسی شد. سه پیکربندی نخست تنها از یکی از توابع زیان L_1 فروبنیوس یا کوشی استفاده می‌کنند. در سه پیکربندی بعدی، ترکیب‌های دوتایی و درنهایت، پیکربندی کامل شامل هر سه تابع زیان $L_1 + Fro + Ca$ مورد ارزیابی قرارگرفته است. تمامی آزمایش‌ها بر روی مجموعه‌داده Fashion و تحت سه سناریوی نوفه گاوسی، لاپلاسی و کوشی انجام‌شده و عملکرد مدل با استفاده از معیارهای ACC و NMI اندازه‌گیری شده است. نتایج در جدول ۸ ارائه‌شده‌اند؛ بهترین و دومین بهترین مقادیر در هر ستون به‌ترتیب با حروف پررنگ و حروف زیرخط‌دار نمایش داده‌شده‌اند.

برای تحلیل دقیق‌تر نقش مجموعه ابهام و تبیین چرایی عملکرد متفاوت توابع زیان در شرایط نوفه‌دار مختلف، لازم است ارتباط میان ویژگی آماری نوفه و شکل تابع زیان موردتوجه قرار گیرد. در چارچوب استوار توزیعی، هر تابع زیان متناظر با یک فرض مشخص درباره رفتار آماری خطای بازسازی است؛ بااین‌حال، ازآنجاکه توزیع واقعی نوفه در داده‌ها معمولاً از پیش مشخص نیست، اتکا به یک تابع زیان منفرد می‌تواند عملکرد روش را به یک فرض خاص درباره ساختار نوفه وابسته کند. ازاین‌رو، استفاده هم‌زمان از توابع زیان با ویژگی‌های آماری متفاوت، امکان پوشش طیف گسترده‌تری از رفتارهای نوفه‌ای را فراهم می‌کند.

به‌طور مشخص، تابع زیان فروبنیوس که متناظر با خطای مربعی است، با فرض نوفه گاوسی سازگاری بیشتری دارد؛ زیرا در این

حالت، خطاهای کوچک و بزرگ بر اساس مربع مقدار خطا در تابع هدف اثر می‌گذارند و این ساختار با مدل‌سازی خطاهای با توزیع گاوسی مرتبط است. در مقابل، تابع زیان L_1 با رشد خطی نسبت به قدرمطلق خطا، حساسیت کمتری نسبت به خطاهای بزرگ دارد و ازاین‌رو برای شرایطی که خطاهای بزرگ و پراکنده، مشابه رفتار نوفه لاپلاسی، محتمل‌تر هستند، مناسب‌تر است. تابع زیان کوشی نیز به دلیل رشد کندتر نسبت به خطاهای بسیار بزرگ، در مواجهه با توزیع‌های دارای دنباله‌های سنگین و انحراف‌های شدید، استواری بیشتری ایجاد می‌کند. بنابراین، این سه تابع زیان را می‌توان به‌ترتیب به‌عنوان دیدگاه‌های متفاوتی نسبت به خطاهای با رفتار گاوسی، لاپلاسی و دنباله‌سنگین در نظر گرفت.

بااین‌حال، این ارتباط به معنای آن نیست که در حضور یک نوع نوفه مشخص، تنها تابع زیان متناظر با آن توزیع باید عملکرد مطلوبی داشته باشد. در مسئله NMF، تابع زیان مستقیماً در فرآیند یادگیری عوامل نهفته و استخراج ساختار داده دخالت دارد و عملکرد نهایی علاوه بر توزیع نوفه، به‌شدت و ساختار آلودگی، داده مورداستفاده و تعامل میان اجزای مختلف مدل نیز وابسته است. ازاین‌رو، انتظار می‌رود ترکیب توابع زیان مکمل بتواند در شرایطی که توزیع واقعی نوفه نامعلوم، ترکیبی یا با فرض اولیه متفاوت باشد، انعطاف‌پذیری بیشتری نسبت به یک تابع زیان منفرد ایجاد کند.

بررسی تجربی این موضوع در جدول ۸ و مقایسه این پیکربندی‌ها امکان بررسی دو موضوع را فراهم می‌کند: نخست، اینکه آیا هر تابع زیان در شرایط نوفه‌دار متناظر با ویژگی‌های آماری خود مزیت نسبی دارد؛ و دوم، اینکه آیا گسترش مجموعه ابهام و فراهم کردن امکان انتخاب تطبیقی میان توابع زیان می‌تواند عملکرد مدل را در برابر شرایط نوفه‌ای مختلف بهبود دهد.

جدول (۸): تحلیل اثر مجموعه ابهام و توابع زیان متناظر با توزیع‌های مختلف نوفه در DRNMF-bS

	Gaussian		Laplacian		Cauchy	
	Acc	NMI	Acc	NMI	Acc	NMI
L_1	0.5908	0.5813	0.5944	0.5855	0.5600	0.5447
Fro	0.6057	0.5905	0.5812	0.5604	0.5532	0.5301
Ca	0.6313	0.6005	0.6027	0.5846	0.5714	0.5497
L_1+Fro	0.6211	0.6117	0.6180	0.5963	0.5841	0.5733
L_1+Ca	0.6354	0.6032	0.6235	0.6030	**0.6255**	0.5844
Fro+Ca	0.6528	0.6089	0.6133	0.5921	0.5902	0.5806
L_1+Fro+Ca	**0.6854**	**0.6438**	**0.6512**	**0.6330**	0.6191	**0.6164**

نتایج جدول ۸ نشان می‌دهد که هیچ تابع زیان منفردی در تمام شرایط نوفه‌ای به‌طور یکنواخت بهترین عملکرد را ندارد و عملکرد مطلوب‌تر در بسیاری از موارد با ترکیب توابع زیان حاصل می‌شود. به‌ویژه، با گسترش مجموعه ابهام از یک تابع زیان به ترکیب‌های دوتایی و سپس مجموعه کامل شامل L_1، فروبنیوس و کوشی، عملکرد کلی مدل در سناریوهای مختلف نوفه بهبود می‌یابد. برای نمونه، در شرایط نوفه گاوسی، ترکیب‌هایی که تابع فروبنیوس را دربر می‌گیرند عملکرد رقابتی دارند، درحالی‌که در شرایط نوفه لاپلاسی، حضور تابع L_1 در ترکیب‌ها عموماً مفیدتر است. همچنین، در شرایط نوفه کوشی، افزودن تابع زیان کوشی به مجموعه ابهام به بهبود عملکرد کمک می‌کند. این رفتار با ویژگی‌های آماری توابع زیان سازگار است و نشان می‌دهد که هر تابع زیان بخشی متفاوت از فضای عدم‌قطعیت مربوط به نوفه را پوشش می‌دهد.

درنتیجه، مزیت مجموعه ابهام پیشنهادی صرفاً از افزایش تعداد توابع زیان ناشی نمی‌شود، بلکه از تنوع فرضیه‌های مربوط به ساختار نوفه و امکان وزن‌دهی تطبیقی آن‌ها ناشی می‌شود. به‌بیان‌دیگر، زمانی که نوع نوفه از پیش به‌طور قطعی مشخص نیست، مدل مجبور نیست خود را به یک فرض آماری خاص محدود کند؛ بلکه می‌تواند اهمیت نسبی توابع زیان موجود درمجموعه ابهام را متناسب با شرایط داده تعیین کند. این ویژگی، مبنای اصلی استفاده از چارچوب استوار توزیعی درروش پیشنهادی

و یکی از دلایل اصلی برتری مجموعه کامل توابع زیان در مقایسه با پیکربندی‌های تک‌تابعی و دوتایی است.

درمجموع، این آزمایش نشان می‌دهد که عملکرد DRNMF-bS صرفاً ناشی از استفاده از یک تابع زیان استوار خاص نیست. در عوض، تشکیل یک مجموعه ابهام متشکل از توابع زیان مکمل، فضای وسیع‌تری از مدل‌های احتمالی نوفه را در اختیار چارچوب استوار توزیعی قرار می‌دهد و سازوکار وزن‌دهی تطبیقی می‌تواند اهمیت نسبی این توابع را متناسب با شرایط داده تعیین کند. بنابراین، نتایج این مطالعه حذف مؤلفه، در کنار آزمایش حذف مؤلفه مربوط به اجزای اصلی مدل، شواهد تجربی بیشتری در حمایت از نقش مستقل و مکمل مؤلفه‌های پیشنهادی DRNMF-bS فراهم می‌کند.

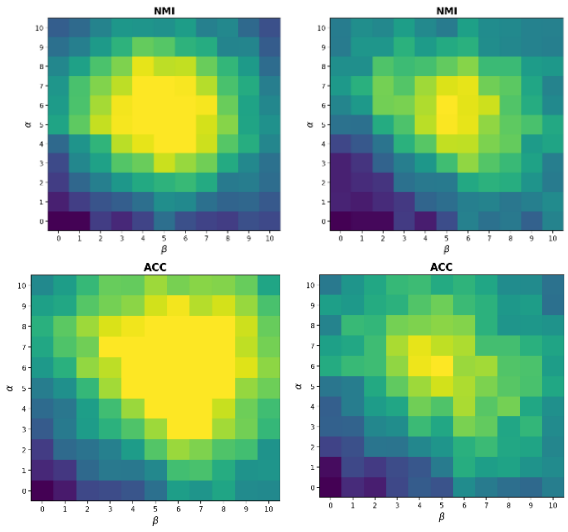
٤-٧ تحلیل حساسیت پارامترهای تنظیمی

به‌منظور بررسی حساسیت عملکرد روش پیشنهادی نسبت به پارامترهای تنظیمی α و β ، آزمایش‌های تکمیلی بر روی دو مجموعه‌داده Blood و USPS انجام شد. برای این منظور، عملکرد روش برای ترکیب‌های مختلف α و β در بازه صفر تا ۱۰ بررسی و با استفاده از دو معیار NMI و ACC ارزیابی شد. نتایج این آزمایش در شکل ٤ ارائه‌شده است. در این شکل، رنگ‌های روشن‌تر نشان‌دهنده مقادیر بالاتر معیار ارزیابی و رنگ‌های تیره‌تر نشان‌دهنده مقادیر پایین‌تر هستند. ازآنجاکه مقادیر هر زیرشکل به‌صورت مستقل در بازه کمینه تا بیشینه نرمال شده‌اند، مقایسه شدت رنگ میان زیرشکل‌های مختلف مدنظر نیست و تحلیل بر اساس تغییرات عملکرد در هر زیرشکل انجام می‌شود.

همان‌گونه که در شکل ٤ مشاهده می‌شود، در هر دو مجموعه‌داده و برای هر دو معیار NMI و ACC ، مقادیر بسیار کوچک α و β عموماً با عملکرد ضعیف‌تری همراه هستند. با افزایش این پارامترها از مقادیر نزدیک به صفر، عملکرد روش بهبودیافته و یک ناحیه با عملکرد بالاتر در مقادیر میانی α و β شکل می‌گیرد. بااین‌حال، افزایش بیش‌ازحد پارامترها نیز بهبود مستمری ایجاد نمی‌کند و در برخی نواحی با کاهش نسبی عملکرد همراه است. بنابراین، انتخاب مقادیر میانی برای α و β ، در مقایسه با مقادیر بسیار کوچک یا بسیار بزرگ، عموماً مناسب‌تر است.

همچنین، ناحیه عملکرد بالاتر در شکل ٤ به یک ترکیب منفرد و بسیار باریک از α و β محدود نیست، بلکه در اطراف مقادیر میانی، ناحیه‌ای نسبتاً گسترده از ترکیب‌های مناسب مشاهده می‌شود. این ویژگی به‌ویژه درمجموعه‌داده USPS مشهود است، درحالی‌که درمجموعه‌داده Blood ناحیه عملکرد بالا نسبتاً متمرکزتر است. این مشاهده نشان می‌دهد که اگرچه عملکرد روش به مقادیر α و β وابسته است، این وابستگی در اطراف ناحیه مناسب شکننده نیست و تغییرات محدود در این پارامترها لزوماً به افت شدید عملکرد منجر نمی‌شود. از سوی دیگر، هم‌پوشانی نسبی نواحی با عملکرد بالاتر در نقشه‌های NMI و ACC نشان می‌دهد که انتخاب مناسب αو β ، در هر دو معیار به شکل هم‌زمان به عملکرد مطلوب‌تر روش منجر می‌شود و نیازی نیست تا برای هر معیار مقادیر متفاوت برای هر پارامتر در نظر گرفته شود.

نکته قابل‌توجه دیگر، عملکرد ضعیف‌تر در نواحی متناظر با مقادیر صفر یا بسیار کوچک α و β است. هنگامی‌که هر دو پارامتر صفر هستند، اثر مؤلفه‌های تنظیمی متناظر در تابع هدف عملاً از بین می‌رود و هنگامی‌که یکی از پارامترها صفر یا بسیار کوچک باشد، سهم مؤلفه متناظر به‌شدت تضعیف می‌شود.



الف ب

شکل (٤): تحلیل حساسیت عملکرد روش پیشنهادی نسبت به پارامترهای تنظیمی α و β بر روی مجموعه‌داده‌های الف) Blood ب) USPS

همان‌طور که شکل ٤ نشان می‌دهد، این حالت‌ها عمدتاً با نواحی تیره‌تر و درنتیجه عملکرد پایین‌تر همراه هستند. این مشاهده ازنظر مفهومی با نتایج مطالعه حذف مؤلفه‌ها در بخش قبل هم‌راستا است؛ در مطالعه حذف مؤلفه‌ها نیز کاهش یا حذف مؤلفه‌های کلیدی روش پیشنهادی موجب افت عملکرد شده بود. بنابراین، نتایج شکل ٤ را می‌توان به‌عنوان شواهد مکملی در تأیید اهمیت مؤلفه‌های تنظیمی پیشنهادی در کنار نتایج مطالعه حذف مؤلفه‌ها در نظر گرفت.

٤-٨ تحلیل همگرایی

به‌منظور بررسی دقیق‌تر رفتار همگرایی مدل پیشنهادی DRNMF-bS، روند تغییرات مقدار تابع هدف (یا نرخ خطا) برحسب تعداد تکرارها بر روی دو مجموعه‌داده Yale و UMIST طی ۳۰۰ تکرار محاسبه و در شکل ۵ ارائه‌شده است. منحنی‌های حاصل برای هر دو مجموعه‌داده از الگوی رفتاری مشابهی پیروی می‌کنند.

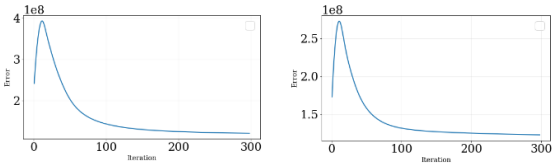
مطابق شکل ۵، مقدار خطا در تکرارهای ابتدایی تا حدود ۳۰ تکرار افزایش ملایمی را تجربه می‌کند و پس‌ازآن به‌صورت یکنواخت کاهش‌یافته و درنهایت به یک مقدار پایدار همگرا می‌شود. این رفتار را می‌توان با توجه به ساختار بهینه‌سازی مدل به‌صورت تحلیلی تبیین کرد.

مدل DRNMF-bS یک مسأله بهینه‌سازی غیرمحدب و چندهدفه را حل می‌کند که در آن، متغیرهای نمایش نهفته و ضرایب وزن‌دهی زیان‌ها(λ) به‌صورت تناوبی به‌روزرسانی می‌شوند. در هر مرحله، با ثابت در نظر گرفتن سایر متغیرها، زیرمسأله مربوطه حل می‌شود که منجر به کاهش مقدار تابع هدف نسبت به همان بلوک متغیرها می‌گردد. بااین‌حال، ازآنجاکه ضرایب λ به‌صورت تطبیقی و وابسته به ساختار داده و الگوهای نوفه‌ای به‌روزرسانی می‌شوند، مقدار تابع هدف کلی در تکرارهای ابتدایی ممکن است دچار نوسان کوتاه‌مدت شود.

به‌عبارت دقیق‌تر، در مراحل اولیه متغیرهای فضای نهان در نزدیکی مقادیر تصادفی اولیه هستند و ضرایب λ هنوز به نزدیکی نقطه تعادل بهینه نرسیده‌اند و مدل در حال جست‌وجوی متغیرهای فضای نهان و توزیع مناسب وزن‌ها میان مؤلفه‌های مختلف زیان است. این بازتنظیم تطبیقی می‌تواند موجب افزایش موقتی مقدار خطا شود، زیرا سهم نسبی هر مؤلفه زیان در حال بازتعریف است. مدل در این فاز، فضای جست‌وجوی خود را به سمت نواحی سازگارتر با توزیع واقعی نوفه هدایت می‌کند؛ فرآیندی که می‌تواند با نوسانات کوتاه‌مدت همراه باشد.

پس از تثبیت نسبی ضرایب تطبیقی (تقریباً پس از ۳۰ تکرار)، به‌روزرسانی‌های تناوبی وارد فاز پایدار می‌شوند. در این مرحله، هر گام به‌روزرسانی موجب کاهش یکنواخت مقدار تابع هدف شده و الگوریتم به سمت یک نقطه ایستا همگرا می‌شود. ازآنجاکه تابع هدف تحت قیود نامنفی بودن و منظم‌سازی bi-stochastic کراندار از پایین است، و هر زیرمسأله نیز کاهش موضعی ایجاد می‌کند، توالی مقادیر تابع هدف در فاز اصلی الگوریتم نزولی و کراندار بوده و درنتیجه همگرا است.

نکته حائز اهمیت آن است که رفتار مشابه مشاهده‌شده در هر دو مجموعه‌داده Yale و UMIST نشان می‌دهد که این الگوی همگرایی وابسته به یک داده خاص نبوده و ناشی از سازوکار درونی مدل است. بنابراین، افزایش موقتی اولیه را می‌توان به‌عنوان مرحله تنظیم تطبیقی وزن‌ها تفسیر کرد، درحالی‌که کاهش یکنواخت پس‌ازآن، بیانگر پایداری الگوریتم و همگرایی به یک جواب استوار و ساختاریافته است.



الف ب

شکل (۵): نمودار همگرایی خطای بازسازی الف) Yale ب) UMIST

٤-٩ تحلیل زمان اجرای الگوریتم

در این بخش به‌منظور ارزیابی هزینه محاسباتی عملی روش پیشنهادی و تکمیل تحلیل پیچیدگی نظری ارائه‌شده در بخش ۳-۵، زمان اجرای DRNMF-bS با روش‌های مقایسه‌ای بر روی هر ۱۲ مجموعه‌داده اندازه‌گیری و برحسب ثانیه گزارش‌شده است. نتایج در جدول ۹ ارائه‌شده‌اند. مطابق جدول، زمان اجرای روش پیشنهادی درمجموعه‌داده‌های مختلف متناسب با اندازه و پیچیدگی داده تغییر می‌کند؛ به‌طوری‌که برای مجموعه‌داده‌های کوچک مانند

Seeds و Ecoli به‌ترتیب تنها ۰/۲۱ و ۰/۳۶ ثانیه و برای مجموعه‌داده‌های بزرگ‌تر مانند OrganA و Blood به‌ترتیب ۴۲/۷ و ۳۸/۶ ثانیه است. این روند با تحلیل پیچیدگی محاسباتی روش سازگار است، زیرا علاوه بر عملیات مربوط به به‌روزرسانی عوامل NMF و وزن‌های نمونه، به‌روزرسانی ماتریس شباهت دوتصادفی نیز در هر تکرار انجام می‌شود. بنابراین، با افزایش تعداد نمونه‌ها، هزینه مربوط به مؤلفه گراف نیز افزایش‌یافته و درمجموعه‌داده‌های بزرگ‌تر سهم بیشتری از زمان اجرا را به خود اختصاص می‌دهد. از منظر مقایسه‌ای، اگرچه DRNMF-bS به دلیل ترکیب یادگیری استوار توزیعی با یادگیری تطبیقی گراف، زمان اجرای بیشتری نسبت به روش‌های پایه دارد، اما به‌عنوان‌مثال به شکل چشمگیری سریع‌تر از DRNMF اجراشده است.

جدول (۹): مقایسه زمان اجرای الگوریتم پیشنهادی با سایر روش‌های مقایسه‌ای بر روی ۱۲ مجموعه داده (ثانیه)

Dataset	Frobenius	$L_{2,1}$	Cauchy	EWNMF	rCIM	Huber	Elastic	DANMF	DRNMF	SPLNMF	DRNMF-bS
Yale	0.129	0.162	0.161	0.168	0.178	0.589	0.163	1.81	0.601	0.338	0.216
ORL	0.745	0.931	0.824	0.779	0.934	1.82	1.01	4.46	3.20	1.09	1.97
COIL20	1.80	1.93	1.92	1.95	2.14	2.89	2.26	5.73	7.29	2.41	3.05
MNIST	0.507	0.597	0.563	0.648	0.743	1.08	0.702	3.78	2.31	0.98	0.93
Fashion	0.420	0.611	0.529	0.669	0.741	1.05	0.728	2.86	2.29	0.94	1.12
Seeds	0.052	0.070	0.068	0.074	0.081	0.245	0.076	0.65	0.49	0.19	0.21
Ecoli	0.085	0.112	0.111	0.118	0.127	0.356	0.124	0.92	0.68	0.29	0.36
USPS	0.332	0.421	0.405	0.438	0.481	0.985	0.512	2.46	1.86	0.74	0.76
UMIST	0.266	0.347	0.341	0.354	0.389	0.841	0.405	2.01	1.44	0.62	0.65
OrganA	15.50	20.55	21.54	20.65	25.19	35.21	25.15	44.9	82.2	32.5	42.7
Blood	12.42	17.25	18.56	17.53	21.61	31.08	20.21	38.7	69.7	27.1	38.6
Pneumonia	2.38	3.10	3.05	3.18	3.45	6.89	3.55	11.8	9.62	4.72	5.16

۵- جمع‌بندی و کارهای آینده

نتایج این پژوهش نشان داد که ترکیب استواری توزیعی با منظم‌ساز دوتصادفی می‌تواند مقاومت و کیفیت نمایش نهفته را به‌صورت مکمل بهبود دهد. تحلیل نتایج نشان می‌دهد که مزیت روش پیشنهادی صرفاً ناشی از استفاده هم‌زمان از چند تابع زیان نیست؛ بلکه یادگیری تطبیقی وزن‌های λ امکان سازگاری سهم توابع زیان با وضعیت خطای بازسازی را فراهم می‌کند، درحالی‌که منظم‌ساز دوتصادفی با تقویت انسجام روابط میان نمونه‌ها، ساختار نمایش را بهبود می‌دهد. نتایج مطالعه حذف مؤلفه‌ها و مقایسه مجموعه‌های مختلف توابع زیان نیز این تفسیر را تأیید می‌کنند و نشان می‌دهند که عملکرد مطلوب حاصل تعامل این دو سازوکار است. پایداری عملکرد در شرایط مختلف نوفه، به‌ویژه در مقایسه با روش‌هایی که به یک مدل نوفه مشخص متکی هستند، نشان‌دهنده مزیت اصلی چارچوب پیشنهادی در مواجهه با عدم‌قطعیت نوع نوفه است.

در مقابل، این بهبود با افزایش هزینه محاسباتی همراه است که عمدتاً ناشی از یادگیری ساختار دوتصادفی است و در داده‌های بزرگ‌تر می‌تواند به یک چالش مقیاس‌پذیری تبدیل شود. همچنین، عملکرد روش همچنان به انتخاب مناسب مجموعه توابع زیان و پارامترهای تنظیمی وابسته است. بنابراین، روش پیشنهادی را می‌توان رویکردی دانست که در ازای هزینه محاسباتی بیشتر، مقاومت بالاتر و انعطاف‌پذیری بیشتری در یادگیری نمایش فراهم می‌کند؛ سازشی که در نتایج تجربی این پژوهش عموماً به نفع روش پیشنهادی بوده است.

در پژوهش‌های آینده، توسعه مجموعه ابهام به‌صورت داده‌محور و قابل انطباق با الگوهای پیچیده‌تر نوفه، طراحی نسخه‌های مقیاس‌پذیرتر برای یادگیری گراف، و تعمیم چارچوب به داده‌های چندنما و چندوجهی می‌تواند موردبررسی قرار گیرد. همچنین، ارزیابی روش در شرایط واقع‌بینانه‌تری مانند نوفه‌های ترکیبی و ناهمگن، داده‌های ناقص و تغییر توزیع، در کنار مطالعه دقیق‌تر اثر

مقداردهی اولیه و مسیر بهینه‌سازی بر نقاط ایستای حاصل، از مسیرهای مهم برای ادامه این پژوهش هستند.

References

[1] H. Cai, B. Liu, Y. Xiao, L. Lin, Semi-supervised multi-view clustering based on orthonormality constrained nonnegative matrix factorization, Information Sciences 536 (2020) 171–184.

[2] B. Pan, C. Li, H. Che, Nonconvex low-rank tensor approximation with graph and consistent regularizations for multi-view subspace learning, Neural Networks 161 (2023) 638–658.

[3] S. Mei, W. Zhao, Q. Gao, M. Yang, X. Gao, Joint feature selection and optimal bipartite graph learning for subspace clustering, Neural Networks 164 (2023) 408–418.

[4] M. Campbell, P. Miller, K. Díaz-Chito, X. Hong, N. McLaughlin, F. Parvinzamir, J. Martínez Del Rincón, N. O'Connell, A computer vision approach to monitor activity in commercial broiler chickens using trajectory-based clustering analysis, Computers and Electronics in Agriculture, (2024).

[5] A. Petukhova, J. P. Matos-Carvalho, N. Fachada, Text clustering with large language model embeddings, International Journal of Cognitive Computing in Engineering 6 (2025) 100–108.

[6] M. Hajihosseinlou, A. Maghsoudi, R. Ghezelbash, Intelligent mapping of geochemical anomalies: Adaptation of dbscan and mean-shift clustering approaches, Journal of Geochemical Exploration (2024).

[7] X. Zhou, Q. Yang, Q. Liu, W. Liang, K. Wang, Z. Liu, J. Ma, Q. Jin, Spatial–temporal federated transfer learning with multi-sensor data fusion for cooperative positioning, Information Fusion (2024).

[8] C. Liu, R. Li, S. Wu, H. Che, D. Jiang, Z. Yu, H.-S. Wong, Self-guided partial graph propagation for incomplete multiview clustering, IEEE Transactions on Neural Networks and Learning Systems (2024).

[9] A. M. Ikotun, A. E. Ezugwu, L. Abualigah, B. Abuhaija, J. Heming, K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data, Information Sciences 622 (2023) 178–210.

[10] L. Ding, C. Li, D. Jin, S. Ding, Survey of spectral clustering based on graph theory, Pattern Recognition 151 (2024) 110366.

[11] X. Liu, P. Song, C. Sheng, W. Zhang, Robust multi-view non-negative matrix factorization for clustering, Digital Signal Processing 123 (2022) 103447.

[12] D. D. Lee, H. S. Seung, Learning the parts of objects by non-negative matrix factorization, Nature (1999).

[13] D. D. Lee, H. S. Seung, Algorithms for non-negative matrix factorization, Advances in neural information processing systems 13 (556-562) (2001).

[14] S. Ahmadian, K. Berahmand, M. Rostami, S. Forouzandeh, P. Moradi, M. Jalili, Recommender systems based on nonnegative matrix factorization: A survey, IEEE Transactions on Artificial Intelligence 6 (10) (2025) 2554–2574.

[15] F. Saberi-Movahed, K. Berahmand, R. Sheikhpour, Y. Li, S. Pan, M. Jalili, Nonnegative matrix factorization in dimensionality reduction: A survey, ACM Comput. Surv. (Sep. 2025).

[16] Y.-T. Guo, Q.-Q. Li, C.-S. Liang, The rise of nonnegative matrix factorization: Algorithms and applications, Information Systems 123 (2024) 102379.

[17] W. Wu, S. Kwong, J. Hou, Y. Jia, H. Ip, Simultaneous dimensionality reduction and classification via dual embedding regularized nonnegative matrix factorization, IEEE Trans. Image Process. 28 (8) (2019) pp. 3836–3847.

[18] N. K. Logothetis, D. L. Sheinberg, Visual object recognition, Annu. Rev. Neurosci. 19 (2) (1996) 577–621.

[19] Y. Ru, M. Gruninger, Y. Dou, Robust self-supervised symmetric nonnegative matrix factorization to the graph clustering, Scientific Reports 15 (1) (2025).

[20] R. Mahmoodi, S. A. Seyedi, F. Akhlaghian Tab, A. Abdollahpouri, Link prediction by adversarial nonnegative matrix factorization, Knowledge-Based Systems 280 (2023) 110998.

[21] M. Faraji, S. A. Seyedi, F. Akhlaghian Tab, R. Mahmoodi, Multi-label feature selection with global and local label correlation, Expert Systems with Applications 246 (2024) 123198.

[22] S. A. Seyedi, F. Akhlaghian Tab, A. Lotfi, N. Salahian, J. Chavoshinejad, Elastic adversarial deep nonnegative matrix factorization for matrix

completion, Information Sciences 621 (2023) 562–579.
[23] S. Shu, T.-Z. Huang, J. Huang, G. Vivone, Nesterov-accelerated non-negative matrix factorization unrolling network for hyperspectral unmixing, Neurocomputing 653 (2025) 131202.
[24] J. Chavoshinejad, S. A. Seyedi, F. Akhlaghian, N. Salahian, "Self- supervised semi-supervised nonnegative matrix factorization for data clustering," Pattern Recognition (2023).
[25] X. Peng, D. Chen, D. Xu, Hyperplane-based nonnegative matrix factorization with label information, Inf. Sci. 493, (2019), pp. 1–9.
[26] C. Ding, T. Li, W. Peng, H. Park, Orthogonal nonnegative matrix t-factorizations for clustering, KDD '06, Association for Computing Machinery, New York, NY, USA, 2006, p. 126–135.
[27] Y. Dong, H. Che, M.-F. Leung, C. Liu, Z. Yan, Centric graph regularized log-norm sparse nonnegative matrix factorization for multi-view clustering, Signal Processing 217 (2024) 109341.
[28] J. Li, G. Zhou, Y. Qiu, Y. Wang, Y. Zhang, S. Xie, Deep graph regularized non-negative matrix factorization for multi-view clustering, Neurocomputing 390 (2020) 108–116.
[29] A. Dache, A. Vandaele, N. Gillis, Orthogonal symmetric nonnegative matrix tri-factorization, in: 2024 IEEE 34th International Workshop on Machine Learning for Signal Processing (MLSP), 2024, pp. 1–6.
[30] A. Ben Hamza and David J. Brady, Reconstruction of Reflectance Spectra Using Robust Nonnegative Matrix Factorization, IEEE TRANSACTIONS ON SIGNAL PROCESSING, VOL. 54, NO. 9, SEPTEMBER 2006.
[31] Qifa Ke and T. Kanade. Robust l/sub1/ norm factorization in the presence of outliers and missing data by alternative convex programming. In 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), volume 1, pages 739–746 vol. 1, 2005.
[32] Deguang Kong, Chris Ding, and Heng Huang. Robust nonnegative matrix factorization using l21-norm. In Proceedings of the 20th ACM International Conference on Information and Knowledge Management, CIKM '11, page 673–682, New York, NY, USA, 2011.
[33] Zechao Li, Jinhui Tang, and Xiaofei He. Robust structured nonnegative matrix factorization for image representation. IEEE Transactions on Neural Networks and Learning Systems, 29(5):1947–1960, 2018.
[34] Antoine Liutkus, Derry Fitzgerald, and Roland Badeau. Cauchy nonnegative matrix factorization. In 2015 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA), pages 1–5, 2015.
[35] Siyuan Peng, Wee Ser, Zhiping Lin, and Badong Chen. Robust sparse nonnegative matrix factorization based on maximum correntropy criterion. In 2018 IEEE International Symposium on Circuits and Systems (ISCAS), pages 1–5, 2018.
[36] Hongchang Gao, Feiping Nie, Weidong Cai, and Heng Huang. Robust capped norm nonnegative matrix factorization: Capped norm nmf. In Proceedings of the 24th ACM International on Conference on Information and Knowledge Management, CIKM '15, page 871–880, New York, NY, USA, 2015. Association for Computing Machinery.
[37] Naiyang Guan, Tongliang Liu, Yangmuzi Zhang, Dacheng Tao, and Larry S. Davis. Truncated Cauchy non-negative matrix factorization. IEEE Transactions on Pattern Analysis and Machine Intelligence, 41(1):246–259, 2019.
[38] Herbert E Scarf, KJ Arrow, and S Karlin. A min-max solution of an inventory problem. Technical report, Rand Corporation Santa Monica, 1957.
[39] Nicolas Gillis, Le Thi Khanh Hien, Valentin Leplat, and Vincent Y. F. Tan. Distributionally robust and multi-objective nonnegative matrix factorization. IEEE Transactions on Pattern Analysis and Machine Intelligence, 44(8):4052–4064, 2022.
[40] Wafa Barkhoda, Amjad Seyedi, Nicolas Gillis, and Fardin Akhlaghian Tab. Instance-wise distributionally robust nonnegative matrix factorization. Pattern Recognition, 169:111732, 2026.
[41] Wafa Barkhoda, Amjad Seyedi, Nicolas Gillis, and Fardin Akhlaghian Tab. Distributionally robust nonnegative matrix factorization with self-paced adaptive multi-loss fusion. Information Sciences, 728:122823, 2026.
[42] Q. Wang, X. He, X. Jiang, and X. Li, "Robust bi-stochastic graph regularized matrix factorization for data clustering," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 44, no. 1, pp. 390– 403, 2022.

[43] D. Cai, X. He, J. Han, and T. S. Huang, “Graph regularized nonnegative matrix factorization for data representation,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 33, no. 8, pp. 1548–1560, 2011.

[44] X. Wang, F. Nie, and H. Huang, “Structured doubly stochastic matrix for graph based clustering: Structured doubly stochastic matrix,” in Proceedings of the 22nd ACM SIGKDD International conference on Knowledge discovery and data mining. ACM, 2016, pp. 1245–1254.

[45] C. Ding, X. He, and H. D. Simon, “On the equivalence of nonnegative matrix factorization and spectral clustering,” in Proceedings of the 2005 SIAM International Conference on Data Mining. SIAM, 2005, pp. 606–610.

[46] R. Zass and A. Shashua, “Doubly stochastic normalization for spectral clustering,” in Advances in Neural Information Processing Systems, 2007, pp. 1569–1576.

[47] J. Wei, C. Tong, B. Wu, Q. He, S. Qi, Y. Yao, Y. Teng, An entropy weighted nonnegative matrix factorization algorithm for feature representation, IEEE Transactions on Neural Networks and Learning Systems 34 (9) (2023) 5381–5391.

[48] L. Du, X. Li, Y.-D. Shen, Robust nonnegative matrix factorization via half-quadratic minimization, in: 2012 IEEE 12th International Conference on Data Mining, 2012, pp. 201–210.

[49] H. Xiong, D. Kong, Elastic nonnegative matrix factorization, Pattern Recognition 90 (2019).

[50] N. Salahian, F. A. Tab, S. A. Seyedi, J. Chavoshinejad, Deep autoencoder-like NMF with contrastive regularization and feature relationship preservation, Expert Systems with Applications, 214 (2023).

[51] S. Huang, P. Zhao, Y. Ren, T. Li, Z. Xu, Self-paced and soft-weighted nonnegative matrix factorization for data representation, Knowledge-Based Systems 164 (2019) 29–37.

[52] D. Geman, G. Reynolds, Constrained restoration and the recovery of discontinuities, IEEE Transactions on Pattern Analysis and Machine Intelligence 14 (3) (1992) 367–383.

[53] J. von Neumann, Functional Operators, Vol. II: The Geometry of Orthogonal Spaces, Annals of Mathematics Studies, Princeton University Press, 1950.

[54] M. Frank, P. Wolfe, An algorithm for quadratic programming, Naval Research Logistics Quarterly 3 (1-2) (1956) 95–110.

Journal of Applied And Basic Machine Intelligence Research (ABMIR)

Applied and Basic Machine Intelligence Research

Vol. 4, No. 2, Autumn & Winter 2026, pp. 173- 204

Distributionally robust nonnegative matrix factorization for data clustering

Wafa Barkhoda[1], Fardin Akhlaghian Tab[2*]

[1]Ph.D. student, Department of Computer, University of Kurdistan, Sanandaj, Iran

[2]Associate Professor, Department of Computer, University of Kurdistan, Sanandaj, Iran

Article Information

Original Research Paper

Received:
2026 June 26

Accepted:
2026 September 30

Keywords:
Data clustering; Distributionally robust learning; Robust matrix factorization; Nonnegative matrix factorization; bi-stochastic graph regularization

Corresponding Author*:
f.akhlaghian@uok.ac.ir

Abstract

Nonnegative Matrix Factorization (NMF) is a widely used technique for nonnegative data representation learning and clustering. However, most existing NMF methods assume a predefined noise distribution, leading to degraded performance on real-world data contaminated by unknown, heterogeneous, or mixed noise. To address this limitation, we propose a novel framework, termed Distributionally Robust Nonnegative Matrix Factorization with a bi-Stochastic Graph Regularizer (DRNMF-bS). The proposed method is built upon an instance-wise distributionally robust NMF model, where the objective function is formulated as an adaptive combination of multiple loss functions associated with different noise distributions and optimized through a min–max problem over a probabilistic ambiguity set. This formulation enables adaptive learning of loss weights, allowing dynamic adaptation to diverse noise conditions while improving robustness against a broad range of noise distributions and their mixtures. Furthermore, a bi-stochastic graph regularizer is incorporated to exploit the intrinsic geometric structure of the data by jointly learning an adaptive similarity matrix. An efficient iterative optimization algorithm with guaranteed objective descent is developed. Extensive experiments on 12 benchmark datasets against 10 representative NMF-based methods demonstrate superior clustering accuracy, robustness to noise, and generalization performance.

: 10.22034/ABMIR.2026.24826.1262

E-ISSN: 2821-2037 /The Author 2026. Published by Yazd University This is an open access article under the CC BY 4.0 License https://creativecommons.org/licenses/by/4.0/).