



دانشگاه کردستان

دانشکده‌ی مهندسی

گروه مهندسی نرم افزار کامپیوتر

پایان‌نامه‌ی کارشناسی ارشد رشته کامپیوتر گرایش هوش مصنوعی و رباتیکز

عنوان:

انتخاب ویژگی چندبرچسبه با استخراج همبستگی بر حسب سراسری و محلی

پژوهشگر:

محمد فرجی

استاد راهنما:

دکتر فردین اخلاقیان طاب

استاد مشاور:

سید امجد سیدی

شهریور ۱۴۰۲

الله
الرَّحْمَنُ
الرَّحِيمُ



دانشگاه کردستان
دانشکده مهندسی
گروه مهندسی نرم افزار کامپیوتر

پایان نامه کارشناسی ارشد رشته کامپیوتر گرایش هوش مصنوعی و رباتیکز

عنوان:

انتخاب ویژگی چندبرچسبه با استخراج همبستگی برچسب سراسری و محلی

پژوهشگر:

محمد فرجی

استاد راهنما:

دکتر فردین اخلاقیان طاب

استاد مشاور:

سیدا مجد سیدی

شهریور ۱۴۰۲

کلیهی حقوق مادی و معنوی مترتب بر نتایج مطالعات،
ابتكارات و نوآوری‌های ناشی از تحقیق موضوع
این پایان‌نامه متعلق به دانشگاه کردستان است.



تعهد نامه

اینجانب محمد فرجی دانشجوی کارشناسی ارشد رشته مهندسی کامپیوتر، گرایش هوش مصنوعی و رباتیکز دانشگاه کردستان، دانشکده مهندسی گروه مهندسی نرم افزار کامپیوتر تعهد می نمایم که محتوای این پایان نامه نتیجه تلاش و تحقیقات خود بوده و از جایی کپی برداری نشده و به پایان رسانیدن آن نتیجه تلاش و مطالعات مستمر اینجانب و راهنمایی و مشاوره استاد بوده است.

با تقدیم احترام

محمد فرجی

۱۴۰۴ فروردین



با اسمه تعالی

* تعهدنامه دانشجویان تحصیلات تکمیلی دانشگاه کردستان در انجام پایان نامه *

اینجانب محمد فرجی دانشجویی مقطع کارشناسی ارشد رشته مهندسی کامپیوتر گرایش هوش مصنوعی و رباتیک متعدد می شوم:

۱. صداقت، امانتداری و بی طرفی را در انجام پژوهش و انتشار نتایج حاصل از آن رعایت نمایم.
۲. در نگارش نتیجه پژوهش‌های حاصل از موضوع پایان نامه از بازنویسی نوشتۀ‌های دیگران بدون ذکر منبع، بازی با الفاظ، زیاده‌نویسی، کلی‌گویی و جزماندیشی و تصرف گرایی پرهیز نمایم و نتایج پژوهشی خود را در موعد مقرر و با اطلاع استاد راهنمای منتشر نمایم.
۳. تمامی یافته‌های مستخرج از پایان نامه متعلق به دانشگاه کردستان بوده و لازم است در کلیه مقالات مستخرج از آنها نام دانشگاه کردستان را تحت عنوان (دانشجوی دانشگاه کردستان) یا (دانش آموخته دانشگاه کردستان) ذکر نمایم.
۴. در انتشار مقالات نام استاد (استادان) راهنمای و استاد (استادان) مشاور را در لیست مولفین مقاله ذکر نمایم و از آوردن اسمی افرادی که نقش موثری در انجام پژوهش نداشته‌اند، جداً خودداری نمایم.
۵. در بخش سپاسگزاری مقاله، از تمامی افراد و سازمان‌هایی که در اجرای پژوهش مساعدتی مبذول داشته‌اند با ذکر نوع مشارکت تشکر و قدردانی نمایم.
۶. از انتشار همپوشان یا ارسال همزمان یک مقاله به چند مجله و یا ارسال مجدد مقاله چاپ شده به مجلات دیگر خودداری نمایم.
۷. در صورت عدم رعایت موارد مذکور، دانشگاه کردستان مجاز خواهد بود تا برابر مقررات اقدام نماید.

امضاء دانشجو

- دستور العمل نحوه برخورد با موارد تخطی دانشجویان تحصیلات تکمیلی در هنگام انتشار نتایج پژوهش
- ۱- در موارد زیر دانشگاه کردستان با مجله مربوطه مکاتبه و درخواست خارج نمودن مقاله را نموده و موضوع را به محل کار یا تحصیل بعدی دانشجو اطلاع خواهد داد.
 - الف: چاپ مقاله بدون اطلاع و تأیید استادان راهنمای
 - ب: چاپ نتایج حاصل از پژوهش‌های انجام شده در دانشگاه کردستان بدون ذکر نام دانشگاه
 - ۲- در صورت احراز تخلف از سایر موارد درج شده در تعهد نامه دانشجویی، دانشگاه ضمن مکاتبه با مجله مربوطه، حسب مورد تصمیم‌گیری خواهد نمود.



دانشگاه کردستان
دانشکده مهندسی
گروه مهندسی نرم افزار کامپیوتر

پایان نامه کارشناسی ارشد رشته مهندسی کامپیوتر گرایش هوش مصنوعی و رباتیکز

عنوان:

انتخاب ویژگی چندبرجسبه با استخراج همبستگی برچسب سراسری و محلی

پژوهشگر:

محمد فرجی

در تاریخ ۱۴۰۲/۰۶/۲۱ توسط کمیته تخصصی و هیات داوران زیر مورد بررسی قرار گرفت و با درجه عالی به تصویب رسید.

هیات داوران

۱. استاد راهنما

دکتر فردین اخلاقیان طاب

امضاء

دانشیار

مرتبه علمی

نام و نام خانوادگی

استادیار

استادیار

دکتر محسن رمضانی

استادیار

دکتر امجد سیدی

۲. استاد مشاور

استادیار

۳. استاد داور خارجی

استادیار

۴. استاد داور داخلی

استادیار

مهر و امضاء معاون آموزشی و تحصیلات تکمیلی دانشکده

مهر و امضاء مدیر گروه

تّعديم بوجونازين



تقدیر و تشکر

سپاس و ستایش مرخدایی را جل و جلاله که آثار قدرت او بر چهره روز روشن، تابان است و انوار حکمت او در دل شب تار، درفشان. آفریدگاری که خویشتن را به شناساند و درهای علم را بر ما گشود و عمری و فرصتی عطا فرمود تا بدان، بندہی ضعیف خویش را در طریق علم و معرفت بیازماید.

تشکر و سپاس از استاد داشمند و پرمايهام جناب آقای دکتر فردین اخلاقیان طاب که از محضر پروفیض تدریسشان و علم اندوزیشان بهره‌ها برده‌ام.

با امتنان و تشکر بیکران از مساعدت، لطف و دلسوزی‌های بی‌شائبه استاد، دوست و برادر ارجمندم جناب آقای دکتر سیدامجد سیدی که به واقع ریزه‌خوان دانش ایشان بوده و خواهم بود.



چکیده

در حوزه‌های مختلف کاربردی، داده‌های چندبرچسبه با ابعاد بالا بسیار مورد توجه قرار گرفته‌اند که دو چالش مهم شامل: نمونه‌ها با ویژگی‌های بُعد بالا و همچنین تعداد زیادی برچسب را به همراه دارد. در انتخاب ویژگی چندبرچسبه، هدف انتخاب زیرمجموعه‌ای از ویژگی‌ها از یک مجموعه است که برای پیش‌بینی چندین برچسب یا دسته مرتبط با هر نمونه بسیار مناسب می‌باشد. با این حال، اغلب روش‌های موجود طبقه‌بندی چندبرچسبه، مواردی چون وابستگی‌های برچسب و توزیع نامتعادل برچسب را نادیده گرفته‌اند، هر چند که دارای بینش‌های ارزشمند برای طراحی الگوریتم‌های مؤثر انتخاب ویژگی چندبرچسبه باشند. در این پایان‌نامه، یک مدل انتخاب ویژگی پیشنهاد می‌شود که با بهره‌گیری از همبستگی‌های سراسری و محلی برچسب‌ها، ویژگی‌های متمایزکننده در سراسر برچسب‌ها را انتخاب کند. علاوه‌بر این، با بازنمایی ماتریس ویژگی و ماتریس برچسب در یک فضای مشترک نهان، همبستگی‌های موجود بین ویژگی‌ها و برچسب‌ها را بدست می‌آورد. در این فضای مشترک، الگوها یا ارتباطات مشترکی که در سراسر چندین برچسب و ویژگی وجود دارد، مشخص می‌شود.تابع پیشنهادی شامل $\ell = 2, 1, 2$ و یک الگوریتم مبتنی بر تکرار و بهینه‌سازی متناوب است تا ضرایب خلوت مورد نیاز برای انتخاب ویژگی چندبرچسبه را بدست آورد. ارزیابی روش پیشنهادی بر روی شش معیار ارزیابی و دوازده مجموعه داده انجام شده است و نتایج نشان‌دهنده این است که مدل ارائه شده مؤثر، و عملکرد آن از روش‌های مقایسه شده بهتر است.

کلمات کلیدی: انتخاب ویژگی، یادگیری چندبرچسبه، همبستگی برچسب، تجزیه ماتریس نامنفی

فهرست مطالب

عنوان	صفحة
۱ مقدمه	۱
۱.۱ بیان و اهمیت مسئله	۲
۲.۱ ساختار پایان نامه	۴
۲ پیشینه تحقیق	۶
۲.۱ نمادها	۷
۲.۲ کارهای مرتبط	۷
۳ روش پیشنهادی	۱۴
۳.۱ مقدمه	۱۵
۳.۲ فضای نهان مشترک	۱۵
۳.۳ حفظ ساختار محلی	۱۷
۳.۴ همبستگی برچسب سراسری و محلی	۱۸
۳.۵ بهینه سازی	۲۱
۴ آزمایشات	۲۶
۴.۱ مجموعه داده ها	۲۷
۴.۲ معیارهای ارزیابی	۲۸
۴.۳ روش های مقایسه شده	۳۰
۴.۴ نتایج آزمایشات	۳۱
۴.۵ تجزیه و تحلیل پارامترها	۳۴
۴.۶ تحلیل همگرایی	۳۸
۵ نتیجه گیری و کارهای آینده	۴۲
۵.۱ نتیجه گیری	۴۳

۲۰۵ پیشنهادهایی برای تحقیقات آتی ۴۳

۴۵

مراجع



فهرست تصاویر

عنوان	صفحة
شماي کلی روش پيشنهادي (MLFS-GLOCAL)	۱۵
نمودار رادار در شش معیار مختلف	۳۵
نمودار درصد انتخاب ویژگی ها در معیار Micro-F1	۳۶
نمودار درصد انتخاب ویژگی ها در معیار Macro-F1	۳۶
نمودار درصد انتخاب ویژگی ها در معیار Hamming Loss	۳۷
نمودار درصد انتخاب ویژگی ها در معیار Ranking Loss	۳۷
تجزیه و تحلیل پارامترهای α و β در مدل پیشنهادی	۳۹
تجزیه و تحلیل پارامترهای α و β در مدل پیشنهادی	۴۰
همگرایی مدل پیشنهادی (MLFS-GLOCAL)	۴۱

فهرست جداول

عنوان	صفحة
جزئیات مجموعه‌داده‌های دنیای واقعی	۱.۴ ۲۷
ارزیابی مدل بر اساس معیار Micro-F1	۲.۴ ۳۱
ارزیابی مدل بر اساس معیار Macro-F1	۳.۴ ۳۲
ارزیابی مدل بر اساس معیار Average Precision	۴.۴ ۳۲
ارزیابی مدل بر اساس معیار Ranking Loss	۵.۴ ۳۲
ارزیابی مدل بر اساس معیار Hamming Loss	۶.۴ ۳۳
ارزیابی مدل بر اساس معیار Coverage Error	۷.۴ ۳۳





فصل اول

مقدمه

۱۰.۱ بیان و اهمیت مسئله

به دلیل تأثیرات منفی ویژگی‌های با ابعاد بالا مانند افزایش پیچیدگی یادگیری^۲ [۱، ۲]، افزایش تخصیص فضا^۳ [۳] و کاهش کارایی طبقه‌بندها^۴، توجه به روش‌های انتخاب‌ویژگی^۵ روزبه روز بیشتر شده است [۴]. در طول سال‌های اخیر، انتخاب‌ویژگی به یکی از مسائل مهم تبدیل شده است که تنها با انتخاب‌ویژگی‌های مرتبط^۶ و حذف ویژگی‌های تکراری^۷ و نامربوط^۸ [۵، ۶]، می‌توان تعداد ویژگی‌ها را کاهش و کارایی مدل را بهبود داد. انتخاب‌ویژگی در مسائل مختلفی مانند: سیستم‌های توصیه‌گر^۹، حاشیه‌نویسی تصویر و ویدئو^{۱۰} [۵]، پردازش داده‌های ریزآرایه^{۱۱} و ژنومیکس^{۱۲} [۸] به تدریج در حال گسترش است. از آنجایی که اشیاء در دنیای واقعی در عین حال چندین برچسب^{۱۳} دارند، تحقیقات قابل توجهی در مورد انتخاب‌ویژگی در این حوزه‌ها انجام شده است. برای مثال در ژنومیکس، یک زن می‌تواند چندین وظیفه مانند: فنوستز^{۱۴}، تجزیه پروتئین^{۱۵} و انتقال سیگنال^{۱۶} [۹] داشته باشد.

در تحلیل موسیقی^{۱۷}، یک قطعه موسیقی می‌تواند همزمان دارای چند احساس مختلف مانند غم، شادی و ترس باشد [۱۰]. همچنین در دسته‌بندی اخبار یک خبر می‌تواند جزو دسته‌های مختلفی باشد. در یادگیری چندبرچسبه، برخلاف مسائل چندکلاسه^{۱۸}، همبستگی بین برچسب‌ها درنظر گرفته می‌شود. به همین دلیل، علاوه بر فضای ابعاد، استخراج و استفاده از همبستگی برچسب‌ها این مسئله را به یک مسئله NP-hard تبدیل کرده است.

در حال حاضر روش‌های انتخاب‌ویژگی چندبرچسبه بر روی استخراج همبستگی^{۱۹} برچسب-برچسب، وابستگی برچسب-ویژگی و همبستگی ویژگی-ویژگی کار می‌کنند. روش‌های انتخاب‌ویژگی چندبرچسبه به دو دسته انتقال مسئله^{۲۰} و تطبیق الگوریتم^{۲۱} تقسیم می‌شوند [۱۱]. در رویکرد اول، مسئله به داده تکبرچسبه^{۲۲} تبدیل می‌شود و با استفاده از الگوریتم‌های تکبرچسبه، حل می‌شوند. در رویکرد دوم، محققان تعدادی مدل انطباق الگوریتم ارائه داده‌اند که به صورت مستقیم مسائل چندبرچسبه را حل می‌کنند [۱۲، ۱۳]. همبستگی‌های برچسب عمده‌تاً توسط روش‌های انطباق الگوریتم برای روند انتخاب‌ویژگی استخراج می‌شوند. روش‌های تطبیق الگوریتم نسبت به روش‌های انتقال مسئله،

1 High dimentionality features

12 Genomics

2 Learning complexity

13 Multi-Label

3 Heightened space allocation

14 Photosynthesis

4 Classification

15 Protein breakdown

5 Feature selection

16 Signal transmission

6 Relevant

17 Music analysis

7 Redundant

18 Multi-Class

8 Irrelevant

19 Extraction of correlation

9 Recommender systems

20 Problem transformation

10 Video annotation

21 Algorithm adaption

11 Microarray data processing

22 Single-Lable

بهینه‌تر هستند زیرا در تبدیل مسئله همبستگی بین برچسب‌ها درنظر گرفته نمی‌شود. روش تطبیق الگوریتم به سه رویکرد مرتبه‌اول^۱، مرتبه‌دوم^۲ و مرتبه‌بالا^۳ تقسیم می‌شود. در رویکرد مرتبه‌اول همبستگی بین برچسب‌ها درنظر گرفته نمی‌شود و برچسب‌ها مستقل از یکدیگر متصرور می‌شوند. در رویکرد مرتبه‌دوم همبستگی جفت برچسب درنظر گرفته می‌شود. در عمل، یک رتبه‌بندی بین برچسب‌های مرتبط و غیرمرتبط خواهیم داشت. در رویکرد مرتبه‌بالا، همبستگی بین یک زیرمجموعه یا کل مجموعه برچسب‌ها درنظر گرفته می‌شود. در یادگیری چندبرچسبه، همبستگی بین برچسب‌ها می‌توانند داده‌های مهمی را ارائه دهند، برای مثال: اگر برچسب‌های اسکیت‌بُرد و برف وجود داشته باشند، به احتمال زیاد برچسب ورزش‌های زمستانی نیز وجود خواهد داشت. به طور مشابه اگر برچسب‌های دریا و هوای آفتابی وجود داشته باشند، به احتمال زیاد برچسب برف وجود ندارد. از این‌رو استفاده از سطح‌های مختلف همبستگی برچسب‌ها مورد توجه بسیاری از محققان قرار گرفته است [۱۴]. با توجه به مطالب گفته شده، نویسنده‌گان [۱۵-۱۷] به طور گسترده بر روی همبستگی سراسری برچسب‌ها^۴ تمرکز کرده‌اند. در واقع، بعضی از همبستگی‌های برچسب [۱۸، ۱۹] بر روی همبستگی محلی^۵ برچسب‌ها تمرکز کرده‌اند، برای مثال: در فضای اینترنت، کلمه آمازون به فروشگاه اینترنتی آمازون اشاره دارد، در حالی‌که همین کلمه در فضای طبیعت به جنگل آمازون اشاره دارد. اخیراً همبستگی‌های سراسری و محلی برچسب‌ها مورد توجه پژوهشگران بسیاری قرار گرفته است، که اگر به طور همزمان از دو همبستگی سراسری و محلی برچسب استفاده شود عملکرد یا نتایج بسیار بهتر و تأثیرگذارتر است. با توجه به مطالب بیان شده، در این پایان‌نامه، انتخاب ویژگی چندبرچسبه با استفاده از همبستگی سراسری و محلی^۶ برچسب‌ها ارائه شده است. بررسی‌های انجام شده نشان می‌دهد که استفاده از یک فضای نهان بجای ماتریس اصلی نتایج بسیار بهتری را خواهد داشت از این‌رو فضای برچسب‌ها به یک ماتریس نهان^۷ انتقال می‌یابد، در فضای نهان برچسب‌ها، اطلاعات زائد و تکراری ماتریس اصلی حذف می‌شوند و یک ساختار فشرده‌تر و مفید با یک همبستگی ضمنی^۸ بین برچسب‌ها بدست می‌آید. همچنین با انتقال ماتریس ویژگی^۹ به این ماتریس نهان به راحتی الگوهای مشترک بین ویژگی‌های مشابه^{۱۰} و برچسب‌های مشابه^{۱۱} مشخص می‌شوند. در نهایت اطلاعات ویژگی و برچسب در یک فضای مشترک^{۱۲} خواهند بود. اگرچه می‌توان گفت که در این فضا همبستگی ضمنی بین برچسب‌ها وجود دارد، اما می‌تواند بهتر شود. به این‌منظور همبستگی‌های سراسری و محلی برچسب‌ها از ماتریس برچسب استخراج^{۱۳} شده، این کار سبب می‌شود که پیش‌بینی در

1 First-order

7 Latent matrix

2 Second-order

8 Implicit

3 High-order

9 Feature matrix

4 Global correlation

10 Similar features

5 Local correlation

11 Similar labels

6 Multi-Label Feature Selection with Global and Local Label Correlation

12 Shared space

13 Extract

برچسب‌های بسیار مرتبط، مشابه باشند. برای یافتن ویژگی‌های مرتبط، روش پیشنهادی شامل همبستگی‌های برچسب ضمنی و صریح^۱ است و موجب کاهش اثرات منفی اطلاعات برچسب می‌شود. همچنین به منظور حفظ سازگاری^۲ بین فضای اصلی ویژگی‌ها در فضای نهان از منظم‌ساز خمینه^۳ استفاده می‌کنیم. در نهایت از $\ell_{2,1}$ ^۴ به منظور انتخاب ویژگی‌های متمایز و بهبود تفسیرپذیری ماتریس ضرایب بهره می‌بریم. اهداف اصلی روش ارائه شده به صورت زیر خلاصه می‌شود:

- انتخاب ویژگی‌های متمایز با استفاده از همبستگی سراسری و محلی برچسب‌ها.
- انتقال فضای ویژگی و فضای برچسب به یک فضای مشترک و کم‌بعد^۵ جهت استخراج الگوهای بین ویژگی‌ها و برچسب‌ها.
- معرفی یک مدل مبتنی بر تجزیه ماتریس نامنفی^۶ که دارای خاصیت خوشبندی^۷ و تفسیرپذیری^۸ برای انتخاب ویژگی‌های بهینه می‌باشد.
- استفاده از منظم‌ساز گراف^۹ جهت حفظ سازگاری بین فضای اصلی و نهان.^{۱۰}
- استفاده از یک الگوریتم کمینه‌سازی^{۱۱} مؤثر جهت بهینه‌سازی^{۱۲} مدل پیشنهادی.

۲۰.۱ ساختار پایان نامه

این پایان نامه به فصل‌های زیر تقسیم شده است. در فصل اول، مقدمه‌ای بر پایان نامه شامل تعریف مسئله، تاریخچه و اهمیت آن، هدف از پژوهش، مفروضات و روش انجام پژوهش پرداخته شده است. در فصل دوم، توضیح مفاهیم بنیادی^{۱۳} و کارهای مرتبط در این زمینه بیان می‌شود. در فصل سوم، معرفی روش پیشنهادی مبتنی بر تجزیه ماتریس نامنفی [۲۰]^{۱۴} که دارای خصوصیات ذاتی خوبه‌بندی و تفسیرپذیری جهت انتخاب ویژگی‌های مهم می‌باشد. در فصل چهارم، شرایط آزمایش، مجموعه‌داده‌های مورد استفاده، معیارهای ارزیابی بیان شده و سپس نتایج آزمایش‌های روش پیشنهادی ارائه

1. Explicit

7. Interpretability

2. Consistency

8. Graph regularization

3. Manifold regularization

9. Latent structure

4. Low-dimension

10. Optimization

5. Nonnegative Matrix Factorization

11. Minimization

6. Clustering

12. Foundational concepts

می‌شوند. در فصل پنجم، به جمع‌بندی کلی روش ارائه شده در این پایان‌نامه پرداخته و پیشنهادهایی برای توسعه و بهبود مدل ارائه شده برای کارهای آتی، مطرح می‌گردد.





فصل دوم

پیشینه تحقیق

۱۰۲ نمادها

در این قسمت از حروف بزرگ ضخیم^۱ برای نشان دادن ماتریس استفاده می‌کنیم، مانند ماتریس A . برای هر ماتریس $A \in \mathbb{R}^{n \times m}$ نماد A_i به معنای سطر j -ام از آن ماتریس است. متغیرهای عددی^۲ با حروف کوچک به صورت انگلیسی مورب^۳ (یعنی i, j, n و غیره) نشان داده می‌شوند در حالی که بردارها^۴ با حروف کوچک ضخیم نمایش داده می‌شوند (مانند A, x و غیره). $\text{Tr}(A)$ به معنای اثر^۵ ماتریس A و A^\top به معنای ترانهاده^۶ آن است. نُرم فروبنیوس^۷ ماتریس $A \in \mathbb{R}^{m \times n}$ به صورت $\|A\|_F = \sqrt{\sum_{i=1}^m \sum_{j=1}^n A_{ij}^2} = \sqrt{\text{Tr}(A^\top A)}$ تعریف می‌شود، که A_{ij} به مفهوم (i, j) درایه داخلی آن ماتریس است. ماتریس داده به صورت $X \in \mathbb{R}^{n \times d}$ تعریف می‌شود که n بیانگر تعداد نمونه‌ها و d تعداد ویژگی‌ها می‌باشد، همچنین ماتریس برچسب به صورت $Y \in \mathbb{R}^{n \times l}$ تعریف می‌شود که l بیانگر تعداد برچسب‌ها می‌باشد. اگر نمونه i -ام برچسب j -ام را داشته باشد $Y_{i,j} = 1$ در غیر این صورت 0 .

۲۰۲ کارهای مرتبط

در سال‌های اخیر، پژوهش محققان بر روی روش‌های انتخاب ویژگی برای مدیریت داده‌های چندبرچسبه، که در آن هر نمونه با چندین برچسب مرتبط است، افزایش یافته است. تحقیق در مورد انتخاب ویژگی چندبرچسبه به سرعت پیشرفت کرده زیرا داده‌های چندبرچسبه بیشتری مورد استفاده قرار گرفته است. روش‌های موجود انتخاب ویژگی داده‌های چندبرچسبه عمدهاً مبتنی بر روش‌های تئوری اطلاعات^۸ و روش‌های تعبیه شده^۹ هستند. روش‌های تئوری اطلاعات از اطلاعات متقابل یا اطلاعات متقابل شرطی^{۱۰} برای استخراج همبستگی بین هر ویژگی نامزد^{۱۱} و هر برچسب کلاس

1. Bold

2. Scalar

3. Italic

4. Vector

5. Trace

6. Transpose

7. Frobenius norm

8. Information-theoretic

9. Embedded based

10. Conditional

11. Condimate feature

استفاده می‌کنند. به عنوان مثال، رویکرد حداکثر وابستگی^۱ و حداقل افزونگی^۲ (MDMR) [۲۱]، براساس افزایش وابستگی ویژگی، بین ویژگی‌ها و برچسب‌ها و کاهش افزونگی ویژگی بهترین زیر مجموعه ویژگی را بدست می‌آورد. تابع هدف MDMR به صورت زیر می‌باشد:

$$J(f_k) = \sum_{l_i \in L} I(f_k; l_i) - \frac{1}{|S|} \sum_{f_j \in S} \left\{ I(f_k; f_j) - \sum_{l_i \in L} I(f_k; l_i | f_j) \right\} \quad (۱.۲)$$

f_k, f_j و S به ترتیب ویژگی که انتخاب شده، ویژگی کاندید و مجموعه ویژگی‌هایی که قبلاً انتخاب شده‌اند. همچنین $I(f_k; l_i)$ وابستگی ویژگی‌ها می‌باشد و افزونگی بین ویژگی‌های به وسیله‌ی $I(f_k; f_j) - \sum_{l_i \in L} I(f_k; l_i | f_j)$ اندازه‌گیری می‌شود. مشابه این کار، لی و همکاران پژوهشی را تحت عنوان SCLS [۲۲] ارائه دادند که شامل دو عبارت ارتباط ویژگی^۳ و معیار ارتباط مقیاس‌پذیر^۴ است. آنتروپی ویژگی^۵-ام توسط $H(f_k)$ اندازه‌گیری می‌شود.

$$J(f_k) = \sum_{l_i \in L} I(f_k; l_i) - \sum_{f_j \in S} \frac{I(f_k; f_j)}{H(f_k)} \sum_{l_i \in L} I(f_k; l_i) \quad (۲.۲)$$

اخیراً LRFS برای مسائل انتخاب ویژگی داده‌های چندبرچسبه ارائه شده است [۲۳]. در این روش برچسب‌ها به دسته‌های مستقل و وابسته تقسیم می‌شوند. تابع LRFS به شرح فرمول (۳.۲) مطرح گردیده است.

$$\begin{aligned} J(f_k) &= LR(f_k; L) - \frac{1}{|S|} \sum_{f_j \in S} I(f_k; f_j) \\ &= \sum_{l_i \in L} \left\{ \sum_{l_i \neq l_j, l_j \in L} I(f_k; l_j | l_i) - \frac{1}{|S|} \sum_{f_j \in S} I(f_k; f_j) \right\} \end{aligned} \quad (۳.۲)$$

روش‌های مبتنی بر تئوری اطلاعات، ارتباطات متقابل مرتبه‌بالا بین ویژگی‌ها و برچسب‌ها را نادیده می‌گیرند. بنابراین اهمیت هر یک از ویژگی‌ها یا برچسب‌ها یک عامل کلیدی در تأثیرگذاری این رویکردها می‌باشد. در عوض، رویکردهای مبتنی بر انتخاب ویژگی چندبرچسبه بر استفاده از همبستگی‌های برچسب تأکید می‌کنند و از همبستگی‌های برچسب برای انتخاب زیرمجموعه‌ایی از ویژگی‌های فشرده بهره می‌برند. در سال‌های اخیر، چندین روش مختلف انتخاب ویژگی

1 Max-dependency

2 Min-redundancy

3 Feature relevance

4 Scalable relevance evaluation

5 Feature entropy

مبتنی بر تعبیه خلوت^۱ ارائه شده است [۲۵، ۲۴، ۱۳]. به عنوان یکی از روش‌های پایه، نی و همکاران [۲۴]، انتخاب ویژگی

مؤثر و مقاوم^۲ را از طریق کمینه‌سازی نُرم $\ell_{2,1}$ متصل^۳ (RFS) معرفی کردند که به صورت زیر فرموله می‌شود:

$$\min_W \left[\sum_{i=1}^n \|x_i W - y_i\| + \gamma \|W\|_{2,1} \right] = \min_W [\|XW - Y\|_{2,1} + \gamma \|W\|_{2,1}] \quad (4.2)$$

که در آن $W \in \mathbb{R}^{d \times c}$ ماتریس ضرایب^۴ است و ویژگی‌های متمایز^۵ توسط $\|W\|_{2,1}$ انتخاب می‌شود. اگرچه

یک روش انتخاب ویژگی چندکلاسه است، اما چارچوب آن برای موارد چندبرچسبی نیز مناسب است و در روش‌های

متعدد انتخاب ویژگی چندبرچسبی استفاده شده است. به عنوان مثال، زو و همکاران [۶]، مدل RFS را برای انتخاب

ویژگی چندبرچسبی در برچسب‌های گمشده^۶ (MLMLFS) توسعه دادند. این روش بر اساس فرض نمونه‌های مشابه

برچسب‌های مشابهی دارند، از منظم‌ساز گراف^۷ استفاده می‌کند. افزون برآن، مجموعه‌ای از ویژگی‌های متمایز توسط

نُرم $\ell_{2,1}$ و p با محدودیت $1 < p < 0$ انتخاب می‌شوند. محققان اخیراً تلاش کرده‌اند که از جایگزین‌هایی برای

ماتریس برچسب استفاده کنند به این دلیل که یادگیری مدل‌های رگرسیون چندبرچسبی روی یک ماتریس برچسب

دو دویی^۸ چالش برانگیز است. این مطالعات رامی‌توان به دوروش با استفاده از ماتریس برچسب نهان یا ماتریس شبکه برچسب

به عنوان هدف رگرسیون طبقه‌بندی کرد. یک روش شناخته‌شده و معروف که از ماتریس برچسب نهان استفاده می‌کند،

انتخاب ویژگی آگاهانه داده‌های چندبرچسبی (MIFS) [۱۳] است، که از همبستگی‌های برچسب ضمنی برای انتخاب

متمایزترین ویژگی‌ها بهره می‌برد. علاوه بر این، MIFS ماتریس ابعاد پایین برچسب را برای جلوگیری از رشد نمایی در

تعداد کل ویژگی‌ها و برچسب‌ها در نظر می‌گیرد.تابع هدف MIFS به شرح زیر می‌باشد:

$$\min_{W,V,B} \|XW - V\|_F^2 + \alpha \|Y - VB\|_F^2 + \beta \text{Tr}(V^\top LV) + \gamma \|W\|_{2,1} \quad (5.2)$$

ماتریس ضرایب ویژگی‌ها، X ماتریس ویژگی، V ماتریس ضریب برچسب‌ها و B ماتریس نهان می‌باشد. به عبارتی W

دیگر در روش MIFS، ماتریس فضای اصلی برچسب به فضای کاهش یافته انتقال داده می‌شود. این روش، پیشگام

1 Sparse embedded-based methods

2 Robust

3 Joint $\ell_{2,1}$ -norm minimization

4 Coefficient matrix

5 Discriminative feature

6 Missing

7 Graph regularization

8 Binary

برخی از روش‌های پیشرفته انتخاب ویژگی چندبرچسبه توسعه یافته است. در این مدل، ماتریس چندبرچسبه اصلی به دو ماتریس تجزیه می‌شود که حاوی محتویاتی از علائم مختلط است. در نتیجه، تفسیر آن‌ها می‌تواند چالش برانگیز باشد. برایتلی و همکاران^[۲۶] CMFS را ارائه دادند. در این روش، انتخاب ویژگی چندبرچسبه با استفاده از ماتریس تجزیه سه‌گانه^۱ انجام می‌شود. با استفاده از ماتریس تجزیه سه‌گانه و حفظ ساختار محلی داده‌ها، ماتریس‌های ویژگی و برچسب به یک فضای کم‌بعد نگاشت^۲ می‌شود و با افزایش روابط بین این دو ماتریس در فضای کم‌بعد، نسبت به انتخاب ویژگی اقدام می‌کند. تابع هزینه این روش به صورت زیر است:

$$\begin{aligned} \min_{V,L,Q,P,B} & \|X - VLQ\|_F^2 + \alpha \|Y - VPB\|_F^2 + \beta \|L - P\|_F^2 \\ & + \epsilon \text{Tr}(R(VPB)^\top VPB) + \gamma \|Q\|_{2,1}, \\ \text{s.t. } & V, L, Q, P, B \geq 0 \end{aligned} \quad (۶.۲)$$

انتخاب ویژگی با استفاده از فضای مشترک (SCMFS)^[۲۷]، یکی دیگر از روش‌های توسعه MIFS است. در SCMFS یک فضای مشترک از ماتریس ویژگی و ماتریس برچسب استخراج می‌شود و برای انتخاب ویژگی استفاده می‌گردد. همچنین از روش تجزیه ماتریس نامنفی برای افزایش تفسیرپذیری در فرایند انتخاب ویژگی استفاده شده است. در نهایت با اعمال یک رمزگشا^۳ بر روی ماتریس ویژگی و کمینه‌سازی خطای در عبارت رگرسیون، ویژگی‌های بهینه انتخاب می‌شوند. این روش به صورت فرمول (۷.۲) پیشنهاد شده است.

$$\begin{aligned} \min_{V,B,W} & \|XW - V\|_F^2 + \alpha \|X - VQ\|_F^2 + \beta \|Y - VB\|_F^2 + \gamma \|W\|_{2,1}, \\ \text{s.t. } & W, V, Q, B \geq 0 \end{aligned} \quad (۷.۲)$$

اخیراً گائو و همکاران^[۲۸]، مانند MIFS، انتخاب ویژگی چندبرچسبه با استفاده از ساختار به اشتراک گذاشته^۴ را ارائه دادند، با این تفاوت که عبارت رگرسیون (رمزگذار^۵) در MIFS به یک عبارت تجزیه (رمزگشا) تغییر یافت. فرمول این

¹ Joint tri-factorization

² Mapping

³ Decoder

⁴ Shared structure

⁵ Encoder

مقاله مانند SCMFS بر اساس مدل NMF می‌باشد.

$$\begin{aligned} \min_{V,Q,M} & \|X - VQ^\top\|_F^2 + \alpha\|Y - VM\|_F^2 + \beta\text{Tr}(V^\top LV) + \gamma\|Q\|_{2,1}, \\ \text{s.t. } & V, M, Q \geq 0 \end{aligned} \quad (8.2)$$

روش‌های مبتنی بر شبکه‌برچسب‌های^۱ مختلفی در کاربردهای MLFS وجود دارد که با روش‌های مختلف، بین ماتریس برچسب و جایگزین‌های آن ارتباط برقرار می‌کند. به عنوان مثال، هوانگ و همکاران^[۲۹]، یک روش MLFS با منظم‌ساز خمینه و حداکثرسازی وابستگی (MRDM) پیشنهاد کردند. این روش فضای برچسب را از طریق منظم‌سازی خمینه به یک فضای جایگزین تعییه می‌کند. علاوه بر این، از معیار استقلال هیلبرت اشمیت^۲ (HSIC) به عنوان یک منظم‌ساز برای به حداقل رساندن وابستگی بین خمینه تعییه و ماتریس برچسب استفاده می‌کند.

$$\min_{W, Z^\top Z = I} \|XW - Z\|_F^2 + \alpha\text{Tr}(Z^\top LZ) - \beta\text{Tr}(HZZ^\top HYY^\top) + \gamma\|W\|_{2,1} \quad (9.2)$$

فن و همکاران^[۳۰]، یک چارچوب دوگانه خمینه را پیشنهاد کردند که ماتریس ویژگی را به دو ماتریس نهان (خوشه‌بندی و برچسب) تعییه می‌کند. همچنین از یک منظم‌ساز برای افزایش رابطه‌ی بین خمینه‌ی تعییه و فضای برچسب استفاده می‌کند. همبستگی ضمنی بین برچسب‌ها توسط اطلاعات ساختار سراسری و محلی بهره‌برداری می‌شود.

$$\begin{aligned} \min_{W, V, F, Q} & \|X^\top W - V\|_{2,1} + \gamma\text{Tr}(F^\top X L X^\top F) + \beta\|W - FQ\|_F^2 \\ & + \lambda(\text{Tr}[(V - Y)^\top E(V - Y)] + \text{Tr}(V^\top LV)) + \alpha\|W\|_{2,1}, \\ \text{s.t. } & V \geq 0, W^\top W = I, F^\top F = I \end{aligned} \quad (10.2)$$

به‌طور مشابه در^[۳۱]، یک مدل مقاوم با ترکیب انتخاب ویژگی چندبرچسبی با مدل تفکیک کننده‌ی محلی^۳ ارائه شده است. در این روش با استفاده از اطلاعات مدل تفکیک کننده، نمونه‌ها خوشه‌بندی می‌شوند و همبستگی ضمنی بین

1.Pseudo-labels

2.Hilbert–Schmidt Independence Criterion

3.Local discriminant model

برچسب‌ها نیز استخراج می‌شود.

$$\begin{aligned} \min_{W,b,L,P} & \|XW + 1_n b^\top - Y\|_{2,1} + \alpha \|W - LP\|_F^2 + \beta \text{Tr}(L^\top ML) + \gamma \|W\|_{2,1}, \\ \text{s.t. } & L^\top L = I \end{aligned} \quad (11.2)$$

ژانگ و همکاران [۱۲]، یک روشی به نام NMDG یا انتخاب ویژگی چندبرچسبه نامنفی با قید گراف پویا^۱ ارائه کردند. در این روش، یک ماتریس شبه‌برچسب با استفاده از منظم‌ساز خمینه برچسب و رگرسیون خطی آموزش داده می‌شود، سپس با ادغام خمینه ویژگی‌ها و ماتریس شبه‌برچسب، ماتریس گراف پویای لaplacian^۲ ایجاد می‌شود، که برای یادگیری وزن‌های ماتریس ویژگی^۳ استفاده می‌شود.

$$\begin{aligned} \min_{W,b,F} & \|XW + 1_n b^\top - F\|_F^2 + \alpha \text{Tr}(F^\top L_Y F) + \beta \text{Tr}(WL_{F^\top} W^\top) \\ & + \gamma \text{Tr}(W^\top L_{X^\top} W), \text{s.t. } (W, F) \geq 0. \end{aligned} \quad (12.2)$$

علاوه بر روش‌های انتخاب ویژگی چندبرچسبه ذکر شده، روش‌های یادگیری چندبرچسبه خاصی وجود دارند که همبستگی‌های برچسب مرتبه‌بالا را بررسی می‌کنند. به عنوان یک کار پیشگام، ژو و همکاران [۳۲]، یک روش یادگیری چندبرچسبه GLOCAL را ارائه کردند که همزمان برچسب‌های گم‌شده را بازیابی^۴ می‌کند، طبقه‌بند را آموزش می‌دهد و از همبستگی‌های سراسری و محلی برچسب همراه بهینه‌سازی خمینه برچسب و یادگیری یک بازنمایی^۵ برچسب نهان استفاده می‌کند. ژانو و همکاران [۳۳]، روش یادگیری چندبرچسبه به نام LSGL را معرفی کردند، که با درنظر گرفتن سازگاری سراسری و همواری^۶ محلی برچسب‌ها، یک ماتریس همبستگی برچسب را می‌آموزد. LSGL تلاش می‌کند تا همبستگی برچسب را از دیدگاه‌های سراسری و محلی در یک مدل خودبازنمایی^۷ و یک چارچوب ساختار داده محلی استخراج کند. کومار و همکاران [۳۴]، تبدیل زیرفضای برچسب رتبه‌پایین^۸ برای یادگیری چندبرچسبه با برچسب‌های

1 Dynamic graph constraints

2 Laplacian

3 Learning of feature weight matrix

4 Retrive

5 Representation

6 Smoothness

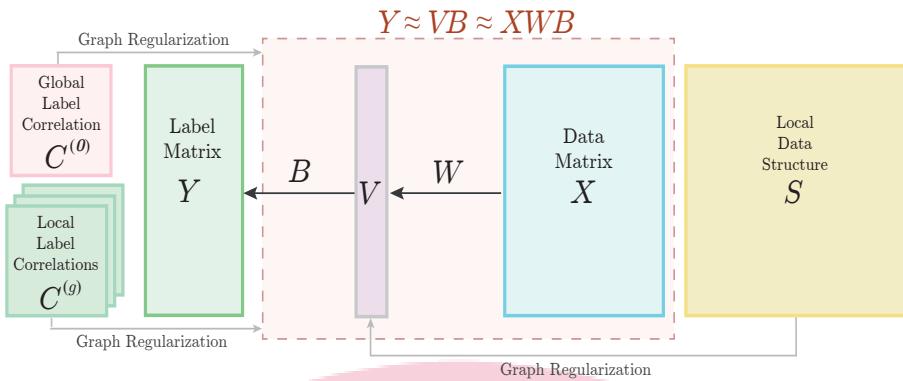
7 Self-representation

8 Low-rank

گم شده LRMML را ارائه کردند. این چارچوب، همبستگی سراسری برچسبها را با یک مدل خودبازنمایی استخراج می‌کند و برای بازیابی برچسبهای گم شده از فضای ماتریس رتبه‌پایین برچسب به همراه آموزش طبقه‌بند استفاده می‌کند.







شکل ۱۰.۳: شمای کلی روش پیشنهادی (MLFS-GLOCAL)

۱.۳ مقدمه

در این فصل، انتخاب ویژگی چندبرچسبه با استفاده از همبستگی سراسری و محلی برچسبها (MLFS-GLOCAL) معرفی می‌شود و از همبستگی برچسب‌های سراسری و محلی برای انتخاب ویژگی‌های مرتبط و غیرتکراری استفاده می‌کند. موفقیت مدل ارائه شده به چهار عامل اصلی بستگی دارد: (۱) برای ارائه یک بازنمایی برچسب-ویژگی بهینه‌تر، ساختار رتبه‌پایین را از ماتریس‌های برچسب و ویژگی استخراج می‌شود. در این فضای کاهش یافته، یک همبستگی ضمنی بین برچسب‌ها وجود دارد (بخش ۲.۳). (۲) جریمه‌ای^۱ اعمال می‌شود که همواری نگاشت محلی را در فضای نهان مشترک حفظ می‌کند (بخش ۳.۳). (۳) با استفاده از اطلاعات همبستگی‌های سراسری و محلی برچسب، از ماتریس برچسب بصورت بهینه‌تر استفاده کرد (بخش ۴.۳). (۴) مسئله، با ترکیب موارد فوق به یک مدل یادگیری مشترک متصل تبدیل می‌شود و از یک رویکرد کمینه‌سازی متناوب مؤثر برای بهینه‌سازی استفاده می‌کند (بخش ۵.۳). شکل ۱۰.۳ نمایش شماتیک از مدل MLFS-GLOCAL را ارائه می‌دهد.

۲.۰.۳ فضای نهان مشترک

ماتریس ویژگی به صورت $Y = \{y_1, \dots, y_l\} \in \mathbb{R}^{n \times d}$ تعریف می‌شود و ماتریس برچسب به صورت $X \in \mathbb{R}^{n \times l}$ تعریف می‌شود که $Y_{i,j} = 1$ اگر نمونه i -ام، برچسب j -ام را داشته باشد، در غیر این صورت $Y_{i,j} = 0$. در مسائل داده‌های چندبرچسبه، برچسب‌ها با یکدیگر مرتبط هستند، از این رو می‌توان از رتبه‌پایین این

¹Penalty

ماتریس استفاده کرد. محتویات ماتریس خلوت Y , دارای مقادیر دودویی و ابعاد بالا می باشد، بنابراین یادگیری براساس این ماتریس مشکل است. لذا اندازه ماتریس کم بعد Y بصورت $k < \min(n, l)$ خواهد بود، از این رو، ماتریس برچسب را به دو ماتریس کوچکتر به صورت زیر تجزیه می کنیم:

$$Y \simeq VB \quad (1.3)$$

که $B \in \mathbb{R}^{k \times l}$ و $V \in \mathbb{R}^{n \times k}$

V ماتریس نهان برچسبها است که فشرده‌تر^۱ و از نظر معنایی انتزاعی‌تر از ماتریس برچسب‌های اصلی هستند. B نشان می‌دهد که چگونه برچسب‌های اصلی با برچسب‌های نهفته مرتبط هستند. اطلاعات مهم ماتریس برچسب اصلی Y در ماتریس کم بعد V کدگذاری شده است که همچنین داده‌های نامطلوب ماتریس برچسب را کاهش می‌دهد. ماتریس‌های V و B با به حداقل رساندن خطای بازنمایی برچسب $\|Y - VB\|_F^2$ بدست می‌آید. بازنمایی برچسب نهان، با V نشان داده شده است، یک ماتریس با ابعاد کم و با مقدار حقيقی که حاوی اطلاعات متراکم می‌باشد. بنابراین یادگیری نگاشت پیوسته از فضای ویژگی به فضای برچسب نهفته نسبتاً آسان‌تر از فضای برچسب اصلی است [۲۲]. یک فرض مهم در رابطه‌ی همبستگی بین ویژگی‌ها و برچسب‌ها این می‌باشد که ویژگی‌های مشابه، برچسب‌های مشابهی دارند. در نتیجه، اطلاعات مشترک بین فضای ویژگی و فضای برچسب باید سازگار باشد. ما در نظر می‌گیریم V یک ماتریس مشترک بین ماتریس ویژگی و ماتریس برچسب باشد. ماتریس $W \in \mathbb{R}^{d \times k}$ بروزرسانی می‌شود جهت نگاشت نمونه‌ها به فضای نهان. با کاهش خطای تابع بازنمایی ویژگی $\|V - XW\|_F^2$ ، ماتریس وزن ویژگی W آموخته خواهد شد و با ادغام بازنمایی برچسب و خطای بازنمایی ویژگی در یک چارچوب ماتریس نامنفی [۲۰]، فضای نهان مشترک را از طریق تابع زیر بدست می‌آوریم:

$$\min_{V, B, W} \|Y - VB\|_F^2 + \|V - XW\|_F^2, \quad \text{s.t.} \quad V, B, W \geq 0 \quad (2.3)$$

¹Compact

۳.۳ حفظ ساختار محلی

در بسیاری از کاربردهای یادگیری ماشین، فضای ویژگی می‌تواند دارای ابعاد بالا همراه با نویز باشد، که استخراج اطلاعات معنی‌دار از داده‌ها را به چالش می‌کشاند. یکی از روش‌های رایج برای استخراج اطلاعات مفید، تبدیل فضای ویژگی اصلی به فضای نهان با ابعاد پایین‌تر است، جایی که ساختار زیربنایی داده‌ها را می‌توان راحت‌تر بدست آورد. با این حال، بازنمایی داده‌ها در فضایی با ابعاد پایین‌تر می‌تواند منجر به از دست رفتن اطلاعات^۱ شود و ممکن است روابط پیچیده بین ویژگی‌ها را به درستی نشان ندهد، از این رو از یک منظم‌ساز گراف [۲۵] برای ارائه یک ماتریس مناسب V که انسجام بین فضای ویژگی اولیه و فضای ساختار نهان را تضمین کند استفاده می‌کنیم. با توجه به اصل اساسی این قاعده‌بندی، اگر دونمونه در فضای ماتریس ویژگی X با یکدیگر همبستگی داشته باشند، در فضای ساختار نهان نیز هر دونمونه v_i و v_j با یکدیگر همبستگی خواهند داشت. به طور خاص، ما از یک منظم‌ساز گراف برای تشویق حفظ ساختار محلی در فضای نهان استفاده می‌کنیم، و اطمینان حاصل می‌کنیم که متغیرهای نهفته به طور دقیق ساختار ویژگی‌های اصلی را منعکس می‌کنند. فرمول منظم‌ساز گراف را می‌توان به صورت زیر بیان کرد:

$$\begin{aligned}
 & \min_V \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \|v_i - v_j\|^2 S_{i,j} \\
 &= \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n (v_i^\top v_i S_{i,j} - 2v_i^\top v_j S_{i,j} + v_j^\top v_j S_{i,j}) \\
 &= \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n (2v_i^\top v_i S_{i,j} - 2v_i^\top v_j S_{i,j}) \\
 &= \sum_{i=1}^n v_i^\top v_i D_{i,i} - \sum_{i=1}^n \sum_{j=1}^n v_i^\top v_j S_{i,j} \\
 &= \text{Tr}(V^\top DV) - \text{Tr}(V^\top SV) = \text{Tr}(V^\top LV),
 \end{aligned} \tag{۳.۳}$$

که D ماتریس قطری^۲ و S ماتریس همبستگی متقارن^۳ است. $L = D - S$ ماتریس لاپلاسین گراف می‌باشد. با ادغام عبارات فوق در مدل، تابع زیر را بدست می‌آوریم:

¹ Loss of information

² Diagonal

³ Symmetric

$$\begin{aligned} \min_{V,B,W} & \|Y - VB\|_F^2 + \|V - XW\|_F^2 + \lambda_1 \text{Tr}(V^\top LV), \\ \text{s.t. } & V, B, W \geq 0 \end{aligned} \quad (4.3)$$

که λ_1 پارامتر حفظ ساختار محلی است.

٤٠٣ همبستگی برچسب سراسری و محلی

برای استفاده مؤثر از اطلاعات برچسب‌ها، باید از همبستگی‌های برچسب استفاده شود. در این روش، مدل پیشنهادی را با استفاده از همبستگی برچسب مرتبه بالا منظم می‌کنیم. در این راستا باید به همزیستی^۱ همبستگی‌های سراسری و محلی برچسب اشاره کرد. برای ترکیب هر دوی آن‌ها، منظم‌ساز خمینه برچسب‌ها را در این قسمت معرفی می‌کنیم. مفهوم منظم‌ساز خمینه سراسری برچسب از منظم‌ساز خمینه سطح نمونه (۳.۳) مشتق شده است. به طور دقیق‌تر اگر دو برچسب دارای همبستگی بالایی باشند، خروجی‌های طبقه‌بند مربوطه آن‌ها نیز باید شباهت داشته باشند و برعکس، برچسب‌هایی که دارای همبستگی کمتری هستند باید خروجی‌های طبقه‌بند آن‌ها شباهت کمتری داشته باشند. به عبارت دیگر، همبستگی‌های برچسب منجر به خروجی‌های طبقه‌بند مشابه می‌شود. در این تحقیق از شباهت کسینوس^۲ برای تعیین کمیت همبستگی سراسری برچسب استفاده می‌شود، که به صورت زیر محاسبه می‌شود:

$$C_{ij} = \frac{y_i^\top y_j}{\|y_i\| \|y_j\|} \quad (5.3)$$

که y_i ، y_j به مفهوم i -ام و j -ام بردار برچسب برای همه نمونه‌ها می‌باشد.

در (۴.۳)، برچسب پیش‌بینی شده برای x به صورت $f(x) = \{f_1, \dots, f_l\}$. $f(x) = xWB$ است، که $f = \{f_1, \dots, f_l\}$. $f(x_n) = \{f(x_1), \dots, f(x_n)\}^\top = XWB$ شامل محتويات پیش‌بینی شده برای i -ام یعنی j -ام برچسب برای نمونه x پیش‌بینی شده است. در نهایت پیش‌بینی برای همه n نمونه در ماتریس پیش‌بینی

¹ Coexistence

² Cosine Similarity

برچسب f_i , اگر برچسب i -ام و j -ام شبیه باشند باید بردار پیش‌بینی شده f_j و f_i این دو برچسب نیز شبیه باشند و بر عکس. تعریف منظم‌ساز خمینه سراسری برچسب شبیه تعریف منظم‌ساز خمینه نمونه (۳۰.۳) است و به صورت زیر تعریف می‌شود:

$$\min_F \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l \|f_i - f_j\|^2 C_{i,j} = \text{Tr}(FAF^\top) - \text{Tr}(FCF^\top) = \text{Tr}(FPF^\top), \quad (6.3)$$

A ماتریس قطری و به صورت $A_{ii} = \sum_{j=1}^l C_{ij}$ تعریف می‌شود. با کمینه‌سازی کردن (۶.۳)، خطای کم خواهد بود. منظم‌ساز خمینه در (۶.۳) به صورت $P = A - C$ ماتریس لایپلاسین است. $F = XWB$ برچسب‌های پیش‌بینی شده برای ویژگی‌ها می‌باشد. از آنجایی که همبستگی‌های برچسب می‌تواند در نواحی مختلف محلی متفاوت باشد، برای استخراج این همبستگی‌ها از منظم‌ساز خمینه محلی استفاده می‌کنیم. فرض می‌کنیم که مجموعه‌داده^۱ X به g گروه $\{X_1, \dots, X_g\}$ تقسیم می‌شود، ماتریس $X_m \in \mathbb{R}^{n_m \times d}$ نمونه دارد. با استفاده از خوشبندی، مانند شبکه‌ها^۲ و مسیرهای ژنی^۳ [۳۶، ۳۷] در کاربردهای بیوانفورماتیک^۴، امکان دستیابی به این پارتیشن‌بندی وجود دارد. فرض $C_m \in \mathbb{R}^{l \times l}$ ماتریس همبستگی برچسب محلی یک گروه m است. و Y_m زیر ماتریس برچسب در Y متناظر با X_m است. مانند (۵.۳)، ما همبستگی‌های برچسب محلی را به صورت زیر محاسبه می‌کنیم:

$$C_{i,j}^{(m)} = \frac{y_i^{(m)\top} y_j^{(m)}}{\|y_i^{(m)}\| \|y_j^{(m)}\|}, \quad m \in \{1, \dots, g\} \quad (7.3)$$

1 Dataset

2 Networks

3 Pathways

4 Bioinformatics

مشابه همبستگی برچسب سراسری (۶.۳)، اگر دو برچسب با همدیگر همبستگی داشته باشند باید بردار پیش‌بینی

شده طبقه‌بند نیز شبیه باشند:

$$\begin{aligned}
 & \min_F \sum_{m=1}^g \frac{n_m}{n} \sum_{i=1}^l \sum_{j=1}^l \|f_i^{(m)} - f_j^{(m)}\|^2 C_{i,j}^{(m)} \\
 &= \sum_{m=1}^g \frac{n_m}{n} [\text{Tr}(F_m A_m F_m^\top) - \text{Tr}(F_m C_m F_m^\top)] \\
 &= \sum_{m=1}^g \frac{n_m}{n} \text{Tr}(F_m P_m F_m^\top)
 \end{aligned} \tag{۸.۳}$$

که $F_m = X_m WB$ ماتریس خروجی طبقه‌بند برای گروه m است و P_m ماتریس لایل‌سین C_m است. برای پوشش عدم تعادل^۱ خوشه، منظم‌ساز همبستگی محلی برچسب را با ضریب n_m/n نرمال می‌کنیم. پس از افزودن منظم‌سازهای خمینه سراسری (۶.۳) و محلی (۸.۳) به تابع (۴.۳)، فرمول زیر بدست می‌آید:

$$\begin{aligned}
 & \min_{V,B,W} \|Y - VB\|_F^2 + \|V - XW\|_F^2 + \lambda_1 \text{Tr}(V^\top LV) \\
 &+ \lambda_2 [\text{Tr}(FPF^\top) + \sum_{m=1}^g \frac{n_m}{n} \text{Tr}(F_m P_m F_m^\top)], \quad \text{s.t. } V, B, W \geq 0
 \end{aligned} \tag{۹.۳}$$

که λ_2 نشان‌دهنده تأثیر همبستگی برچسب سراسری و محلی برای تابع هدف است. در نهایت، تابع ما از نرم $\ell_{2,1}$ ، که برای انتخاب ویژگی استفاده می‌کند. در نتیجه تابع نهایی به صورت زیر تنظیم می‌شود:

$$\begin{aligned}
 & \min_{V,B,W} \|Y - VB\|_F^2 + \|V - XW\|_F^2 + \lambda_1 \text{Tr}(V^\top LV) \\
 &+ \lambda_2 [\text{Tr}(FPF^\top) + \sum_{m=1}^g \frac{n_m}{n} \text{Tr}(F_m P_m F_m^\top)] + \lambda_3 \|W\|_{2,1}, \\
 & \text{s.t. } V, B, W \geq 0
 \end{aligned} \tag{۱۰.۳}$$

^۱ Imbalance

با استفاده از نُرم $\ell_{2,1}$ بر روی سطر ماتریس W خلوتی اعمال می‌کنیم و منظم‌ساز خلوتی تابع هدف نیز توسط پارامتر λ_3 کنترل می‌شود.

۵.۳ بهینه‌سازی

تابع (۱۰.۳) شامل عبارت منظم‌ساز نُرم $\ell_{2,1}$ است. از آنجا که این تابع هموار نیست، همچنین وقتی متغیرهای V و W به صورت ترکیبی در نظر گرفته می‌شوند، غیرمحدب^۱ می‌شود. به عبارت دیگر، ماتریس هسین^۲ که از مشتقات جزئی تابع هدف در درجه دوم بدست می‌آید، ماتریسی نیست که نیمه مثبت معین^۳ باشد. بنابراین، تابع هدف باید به گونه‌ای بهینه‌سازی شود که در هر مرحله از الگوریتم با ثابت نگهداشتمن سه متغیر و به روزرسانی متغیر باقیمانده به صورت محدب^۴ باشد. عبارت $\|W\|_{2,1}$ با استفاده از $\text{Tr}(W^\top DW)$ به شکلی دیگر خفیف می‌شود، که در این حالت D یک ماتریس قطری است [۲۴]. الگوریتم مبتنی بر تکراری ویژگی‌های بروزرسانی در دسترس است. درایه D به شکل $D_{ii} = 1/(\|w_i\| + \epsilon)$, $(\epsilon \leftarrow 0)$ است. که ϵ برای جلوگیری از اختلال در محاسبات مسئله بکار می‌رود. به عبارت دیگر، تابع هدف (۱۰.۳) می‌تواند به صورت زیر بازنویسی شود:

$$\begin{aligned} & \min_{V, B, W} \|Y - VB\|_F^2 + \|V - XW\|_F^2 + \lambda_1 \text{Tr}(V^\top LV) \\ & + \lambda_2 [\text{Tr}(FPF^\top) + \sum_{m=1}^g \frac{n_m}{n} \text{Tr}(F_m P_m F_m^\top)] + \lambda_3 (W^\top DW), \\ & \text{s.t. } V, B, W \geq 0 \end{aligned} \quad (11.3)$$

1 Non-convex

2 Hessian matrix

3 Semi-definite

4 Convex

ما برای افزودن محدودیت نامنفی در تابع، ضرایب لاگرانژی¹ را معرفی می‌کنیم. Φ , Ψ و Ω برای محدود کردن V , B و W

به ترتیب، جایی که $\Omega \in \mathbb{R}^{d \times k}$, $\Psi \in \mathbb{R}^{k \times l}$, $\Phi \in \mathbb{R}^{n \times k}$ در نتیجه، تابع (۱۱.۳) معادل زیر است:

$$\begin{aligned} & \min_{V, B, W} \|Y - VB\|_F^2 + \|V - XW\|_F^2 + \lambda_1 \text{Tr}(V^\top LV) \\ & + \lambda_2 [\text{Tr}(FPF^\top) + \sum_{m=1}^g \frac{n_m}{n} \text{Tr}(F_m P_m F_m^\top)] \\ & + \lambda_3 \text{Tr}(W^\top DW) - \text{Tr}(\Phi V^\top) - \text{Tr}(\Psi B^\top) - \text{Tr}(\Omega W^\top) \end{aligned} \quad (۱۲.۳)$$

با توجه به ماتریس A , $\|A\|_F^2 = \text{Tr}(A^\top A)$. بنابراین، تابع (۱۲.۳) تبدیل می‌شود:

$$\begin{aligned} C = & \text{Tr}[(Y - VB)^\top (Y - VB)] + \text{Tr}[(V - XW)^\top (V - XW)] \\ & + \lambda_1 \text{Tr}(V^\top LV) + \lambda_2 [\text{Tr}(FPF^\top) + \sum_{m=1}^g \frac{n_m}{n} \text{Tr}(F_m P_m F_m^\top)] \\ & + 2\lambda_3 \text{Tr}(W^\top DW) - \text{Tr}(\Phi V^\top) - \text{Tr}(\Psi B^\top) - \text{Tr}(\Omega W^\top) \end{aligned} \quad (۱۳.۳)$$

که V , B و W عبارتند از: $\forall m \{1, \dots, g\}$ ، $F_m = X_m WB$ و $F = XWB$. مشتقات جزئی تابع (۱۳.۳) با توجه به متغیرهای V ، B و W عبارتند از:

$$\begin{aligned} \frac{\partial C}{\partial W} = & -X^\top V + X^\top XW + \lambda_2 [X^\top XWB P B^\top] \\ & + \sum_{m=1}^g \frac{n_m}{n} X_m^\top X_m W B P_m B^\top + 2\lambda_3 DW - \Omega \end{aligned} \quad (۱۴.۳)$$

¹Lagrangian multipliers

$$\begin{aligned} \frac{\partial C}{\partial B} = & -V^\top Y + V^\top VB + \lambda_2 [W^\top X^\top X W B P \\ & + \sum_{m=1}^n \frac{n_m}{n} W^\top X_m^\top X_m W B P_m] - \Phi \end{aligned} \quad (15.3)$$

$$\frac{\partial C}{\partial V} = V - 2XW - 2YB^\top + VBB^\top + \lambda_1 LV - \Psi. \quad (16.3)$$

با قرار دادن مشتقات جزئی (۱۴.۳)، (۱۵.۳) و (۱۶.۳) برابر با صفر، می‌توانیم نقاط بهینه^۱ تابع را پیدا کنیم. با توجه به شرایط^۲ Karush-Kuhn-Tucker را تنظیم می‌کنیم. با حل این

معادلات، توابع بروزرسانی زیر را برای W ، B و V بدست می‌آوریم:

$$\Omega_{SGC} = X^\top X W B C B^\top$$

$$\Omega_{DGC} = X^\top X W B A B^\top$$

$$\Omega_{SLC} = \frac{n_m}{n} X_m^\top X_m W B C_m B^\top$$

$$\Omega_{DLC} = \frac{n_m}{n} X_m^\top X_m W B A_m B^\top$$

$$W \leftarrow W \odot \frac{X^\top V + \lambda_2 [\Omega_{SGC} + \sum_{m=1}^g \Omega_{SLC}]}{X^\top X W + \lambda_2 [\Omega_{DGC} + \sum_{m=1}^g \Omega_{DLC}] + \lambda_3 D W} \quad (17.3)$$

¹ Optimal points

² Conditions

$$\Xi_{SLC} = \frac{n_m}{n} W^\top X_m^\top X_m WBC_m$$

$$\Xi_{DLC} = \frac{n_m}{n} W^\top X_m^\top X_m WBA_m$$

$$\Xi_{SGC} = W^\top X^\top XWBC$$

$$\Xi_{DGC} = W^\top X^\top XWBA$$

$$B \leftarrow B \odot \frac{V^\top Y + \lambda_2 [\Xi_{SGC} + \sum_{m=1}^g \Xi_{SLC}]}{V^\top VB + \lambda_2 [\Xi_{DGC} + \sum_{m=1}^g \Xi_{DLC}]} \quad (18.3)$$

$$V \leftarrow V \odot \frac{XW + YB^\top + \lambda_1 SV}{V + VBB^\top + \lambda_1 DV} \quad (19.3)$$

MLFS-GLOCAL تمام ویژگی‌ها را با استفاده از $\|w_i\|_2$, ($i = 1, \dots, d$) به ترتیب نزولی رتبه‌بندی می‌شود و به ما امکان می‌دهد تا ویژگی‌های با بالاترین امتیاز را بدست آوریم. الگوریتم ۱ مدل MLFS-GLOCAL را به صورت جزئیات تشریح می‌کند.

Algorithm 1 Multi-label Feature Selection with Global and Local label correlation
(MLFS-GLOCAL)

Input: Feature matrix $X \in \mathbb{R}^{n \times d}$ and Label matrix $Y \in \mathbb{R}^{n \times c}$, Regularization parameters λ_1, λ_2 , and λ_3 , and latent factor k ;

Output: Feature score $s_i = \|w_i\|, \forall i \in \{1, 2, \dots, d\}$.

- 1: **Initialize** V, W, B randomly; $t = 0$;
 - 2: **while** $t < \text{MaxIteration}$ **do**
 - 3: Update $D_{ii} \leftarrow \frac{1}{\|w_i\| + \epsilon}$;
 - 4: Update W by (3.17);
 - 5: Update B by (3.18);
 - 6: Update V by (3.19);
 - 7: $t = t + 1$;
 - 8: **end while**
 - 9: **Return** W ;
 - 10: Evaluate the feature score by $s_i = \|w_i\|$.
-



فصل چهارم
آزمایشات

جدول ۱۰.۴ : جزئیات مجموعه‌داده‌های دنیای واقعی

	مجموعه‌داده	# نمونه	# ویرگی	# برچسب
Arts	۵۰۰۰	۴۶۲	۲۶	
Business	۵۰۰۰	۴۳۸	۳۰	
Computers	۵۰۰۰	۶۸۱	۳۲	
corel5k	۵۰۰۰	۴۹۹	۳۷۴	
Education	۵۰۰۰	۵۵۰	۳۲	
Entertainment	۵۰۰۰	۶۴۰	۲۱	
Health	۵۰۰۰	۶۱۲	۳۲	
Recreation	۵۰۰۰	۶۰۶	۲۲	
Reference	۵۰۰۰	۷۹۳	۳۲	
Science	۵۰۰۰	۷۴۳	۴۰	
Social	۵۰۰۰	۱۰۴۷	۳۹	
Society	۵۰۰۰	۶۳۶	۲۷	

این بخش یک ارزیابی جامع برای مدل MLFS-GLOCAL بر روی ۱۲ مجموعه‌داده چندبرچسبه واقعی با استفاده از شش معیار^۱ ارزیابی متنوع انجام می‌دهد. همچنین مدل پیشنهادی با نه روش شناخته‌شده و پیشرفته انتخاب ویژگی مقایسه شده است.

۱۰.۴ مجموعه‌داده‌ها

در بخش آزمایشات، از ۱۲ مجموعه‌داده از کتابخانه Mulan برای طبقه‌بندی متن و تصاویر چندبرچسبه استفاده شده است. مجموعه‌داده‌های چندبرچسبه یاهو^۲ به مجموعه‌ای از چندین مجموعه‌داده اشاره دارد که برای طبقه‌بندی مسائل چندبرچسبه استفاده می‌شوند. این داده‌ها، توسط آزمایشگاه‌های یاهو منتشر شدند و شامل تعداد زیادی اسناد متنی است که به چندین برچسب مرتبط می‌باشند. هر مجموعه‌داده شامل مجموعه آموزش^۳ و مجموعه آزمایش^۴ است که هر کدام به ترتیب شامل ۲۰۰۰ و ۳۰۰۰ سند^۵ هستند [۳۴]. مجموعه‌داده چندبرچسبه Corel5k مجموعه‌ایی از تصاویر است که برای طبقه‌بندی چندبرچسبه استفاده می‌شود. این مجموعه شامل ۵۰۰۰ تصویر از ۳۷۴ دسته‌بندی مختلف است و هر تصویر دارای چندین برچسب است. مجموعه‌داده‌های یاهو و Corel5k به‌طور گسترده در تحقیقات جهت توسعه و ارزیابی الگوریتم‌های طبقه‌بندی چندبرچسبه استفاده شده‌اند. جدول ۱۰.۴ ویژگی‌های هر مجموعه‌داده مرجع

1 Metric

2 Yahoo

3 Train

4 Test

5 Document

را توصیف می‌شود.

۲۰۴ معیارهای ارزیابی

برای ارزیابی عملکرد روش‌های رقیب، از Multi-Label kNN (ML-kNN) [۳۹] برای طبقه‌بندی استفاده می‌شود.

ML-kNN به دلیل تفسیرپذیری و سادگی به عنوان یک الگوریتم متداول برای طبقه‌بندی در رویکردهای انتخاب ویژگی چندبرچسبه استفاده می‌شود [۱۳، ۴۰، ۲۷، ۴۱]. ما $k = 10$ را برای تعداد همسایه‌ها^۱ قرار دادیم، علاوه بر این، از شش معیار ارزیابی متداول استفاده می‌کنیم که شامل Ranking Loss، Average Precision، Macro-F1، Micro-F1 است:

Coverage Error و Hamming Loss هستند. تعاریف این معیارها به شرح زیر است:

: Micro-F1 , Macro-F1 •

هر دو براساس معیار اندازه‌گیری F-measure هستند. معیار ارزیابی مستقیماً از میانگین F-measure استفاده می‌شود برای رتبه‌بندی دقت پیش‌بینی‌ها که توسط طبقه‌بند برچسب تولید شده‌اند.

$$Micro - F1 = \frac{\sum_{i=1}^l 2TP^i}{\sum_{i=1}^l (2TP^i + FP^i + FN^i)}, \quad (1.4)$$

$$Macro - F1 = \sum_{i=1}^l \frac{2TP^i}{2TP^i + FP^i + FN^i}, \quad (2.4)$$

که در آن TP، FP و FN به ترتیب مثبت صحیح^۲، مثبت کاذب^۳ و منفی کاذب^۴ هستند.

:Average Precision •

این معیار مشخص می‌کند درصد برچسب‌هایی که بیشتر از یک برچسب خاص مرتبط هستند:

1. Neighbours

2. True positive

3. False positive

4. False negative

$$AP(D) = \frac{1}{n} \sum_{i=1}^n \frac{1}{\mathbf{1}_m^\top \mathbf{y}_i} \sum_{l:y_i^l=1} \frac{prec_i(l)}{rank_i(l)}, \quad (3.4)$$

$.AP(D) \in [0, 1]$ و $prec_i(l) = \sum_{l:y_i^l=1} \delta(rank_i(l) \geq rank_i(l'))$ که

Ranking Loss • در این معیار، نسبت دو برعکسی که به ترتیب معکوس یا مهم‌تر از برعکس‌های مرتبط در نظر گرفته شده است.

$$RL(D) = \frac{1}{n} \sum_{i=1}^n \frac{1}{\mathbf{1}_m^\top \mathbf{y}_i \mathbf{1}_m^\top \overline{\mathbf{y}}_i} \sum_{l:y_i^l=1} \sum_{l':y_i^{l'}=0} (\delta(rank_i(l) \geq rank_i(l'))), \quad (4.4)$$

که $\overline{\mathbf{y}}_i$ مکمل \mathbf{y}_i در Y و $RL(D) \in [0, 1]$ می‌باشد.

Hamming Loss • در HL در صد برعکس‌هایی را تعیین می‌کند که به اشتباه برعکس گذاری شده‌اند.

$$HL(D) = \frac{1}{n} \sum_{i=1}^n \frac{1}{m} \|h(x_i) \Delta y_i\|_1, \quad (5.4)$$

نماد Δ برای نمایش تفاصل همسان بین دو مجموعه استفاده می‌شود و مجموعه‌ای از مقادیر است که اختصاصاً در یکی از دو مجموعه ظاهر می‌شود.

Coverage Error • در این معیار خطای پوشش یک معیار ارزیابی است که تعداد مراحل یا پیش‌بینی‌های مورد نیاز برای پوشش تمام برعکس‌های مثبت مرتبط با نمونه‌ها را با پایین‌آمدن رتبه‌بندی برعکس‌های پیش‌بینی شده اندازه‌گیری می‌شود.

$$CV(D) = \frac{1}{n} \sum_{i=1}^n \arg \max_{l:y_i^l=1} rank_i(l) - 1, \quad (6.4)$$

مقدار کوچک در Coverage Error، Hamming Loss، Ranking Loss بیان‌گر عملکرد بهتر مدل می‌باشد و مقدار صفر ایده‌آل است، در حالی که در معیارهای Average Precision، Macro-F1، Micro-F1 مقادیر بالا

بيانگر عملکرد بهتر الگوريتم می باشد و مقدار يك مقدار ايده آل می باشد.

٣٠.٤ روش های مقایسه شده

در اين بخش، روش پيشنهادی با روش های جديد انتخاب ويژگی در داده های چندبرچسبه مقایسه شده است که در زير توضيحات مختصری درباره هر کدام از روش ها داده شده است.

- MDMR [٢١]: يك روش انتخاب ويژگی مبتنی بر تئوري اطلاعات است که ويژگی ها را با حداکثر کردن وابستگی و به حداقل رساندن افزاونگی همزمان انتخاب می کند.
- SCLS [٢٢]: يك ديگر از الگوريتم های انتخاب ويژگی است که از تئوري اطلاعات استفاده می کند و ارتباط شرطی را با استفاده از يك معیار ارزیابی ارتباط مقیاس پذیر ارزیابی می کند.
- LRFS [٢٣]: مبتنی بر افزونگی برچسبها می باشد. در اين روش از اطلاعات متقابل شرطی استفاده می شود برای ايجاد يك عبارت جديد ويژگی مرتبط جهت ارزیابی اطلاعات ويژگی ها.
- MIFS [١٩]: يك روش شناخته شده MLFS است که از معنای نهان ماترييس چندبرچسبه برای انتخاب مهم ترین ويژگی ها استفاده می کند.
- CMFS [٢٦]: روشی برای مطالعه اطلاعات ساختار یافته است که بر همبستگی های ويژگی ها و برچسب تکيه دارد.
- SCMFS [٢٧]: از تجزيه ماترييس نامنفي جفت شده برای ايجاد مدل مشترك استفاده می کند.
- SSFS [٢٨]: يك عبارت مشترك ساختار نهان (LSS) را پيشنهاد کرد که هر دو ويژگی نهان و ساختار برچسب را به اشتراك گذاشته و حفظ می کند.
- MRDM [٢٩]: از HSIC به عنوان يك معیار جهت افزایش رابطه بين خمینه تعبيه شده و برچسب های کلاس استفاده می شود.
- NMDG [١٢]: از ماترييس لاپلاسین گراف پويا ساخته شده توسط شبه برچسب در فرآيند انتخاب ويژگی استفاده می کند.

جدول ۲.۴: ارزیابی مدل بر اساس معیار Micro-F1

مجموعه‌داده	SSFS	SCMFS	MIFS	CMFS	MRDM	NMDG	SCLS	LRFS	MDMR	MLFS-GLOCAL
Arts	۰.۲۲۶۳	۰.۲۶۷۴	۰.۱۶۵۵	۰.۱۷۵۳	۰.۲۴۹۶	۰.۱۷۱۴	۰.۱۲۵۸	۰.۰۶۰۱	۰.۱۴۱۸	۰.۳۲۳۰
Business	۰.۶۸۱۶	۰.۶۹۲۷	۰.۶۸۴۶	۰.۶۷۸۶	۰.۶۸۹۲	۰.۶۷۲۸	۰.۶۷۶۰	۰.۶۶۹۸	۰.۶۷۰۴	۰.۶۹۶۶
Computers	۰.۴۲۶۴	۰.۲۲۵۴	۰.۰۳۸۸	۰.۴۱۰۵	۰.۴۲۱۷	۰.۴۰۶۴	۰.۴۳۱۰	۰.۴۰۸۸	۰.۴۰۸۳	۰.۴۵۱۱
Corel5k	۰.۰۲۷۶	۰.۰۳۹۷	۰.۰۳۸۸	۰.۰۳۳۸	۰.۰۳۹۲	۰.۰۳۶۱	۰.۰۳۷۷	۰.۰۱۶۱	۰.۰۲۰۸	۰.۰۴۹۵
Education	۰.۳۰۶۱	۰.۲۱۳۱	۰.۲۰۹۵	۰.۲۸۷۱	۰.۲۳۰۶	۰.۱۱۹۱	۰.۱۳۸۳	۰.۰۹۱۰	۰.۱۵۹۱	۰.۳۵۹۴
Entertainment	۰.۳۲۱۴	۰.۲۶۴۰	۰.۲۷۹۶	۰.۳۰۴۱	۰.۳۵۹۷	۰.۲۲۵۰	۰.۲۶۶۵	۰.۱۲۱۹	۰.۲۸۲۸	۰.۴۳۷۷
Health	۰.۴۸۸۷	۰.۵۱۵۸	۰.۴۶۴۷	۰.۴۷۸۳	۰.۴۹۰۴	۰.۴۶۴۹	۰.۴۳۸۵	۰.۳۶۷۱	۰.۴۱۹۳	۰.۵۷۴۴
Recreation	۰.۲۱۵۸	۰.۲۶۸۹	۰.۲۲۵۲	۰.۱۶۹۴	۰.۲۷۶۰	۰.۲۱۰۴	۰.۱۴۳۲	۰.۰۵۶۹۳	۰.۱۷۶۶	۰.۳۳۷۱
Reference	۰.۴۲۷۱	۰.۴۵۶۳	۰.۴۲۹۱	۰.۴۲۵۹	۰.۴۵۷۰	۰.۳۸۶۴	۰.۴۰۸۳	۰.۳۷۵۲	۰.۳۷۲۲	۰.۴۸۳۸
Science	۰.۲۱۸۵	۰.۲۶۱۵	۰.۱۷۲۵	۰.۱۹۷۵	۰.۲۱۵۳	۰.۱۵۳۰	۰.۱۴۲۲	۰.۰۸۷۱	۰.۱۴۰۶	۰.۲۹۷۸
Social	۰.۰۵۳۰	۰.۰۴۸۵	۰.۴۹۹۵	۰.۵۱۳۲	۰.۵۳۹۸	۰.۴۰۳۵	۰.۴۵۲۰	۰.۳۰۶۹	۰.۴۳۳۰	۰.۵۸۸۳
Society	۰.۳۵۳۴	۰.۲۵۳۶	۰.۳۴۲۹	۰.۳۲۲۴	۰.۳۶۲۵	۰.۳۰۷۱	۰.۲۹۰۴	۰.۲۸۶۲	۰.۲۸۲۳	۰.۳۷۱۱

۴.۴ نتایج آزمایشات

در بخش نتایج آزمایشات، ما ۲۰٪ از ویژگی‌های برتر را برای تعیین میانگین عملکرد برای هر روش انتخاب کردیم. جداول ۲.۴ تا ۷.۴ نتایج این آزمایش‌ها را بر روی شش معیار ارزیابی مختلف نشان می‌دهند. بهترین نتایج برای هر مجموعه‌داده با فونت ضخیم نشان داده می‌شود، جایی که هر چه مقادیر بالاتر باشد، عملکرد طبقه‌بند بهتر است. برای ارائه یک ارزیابی پایدار از عملکرد مدل‌ها، هر روش ۱۰ بار اجرا شده است و میانگین نتایج برای همهٔ مجموعه‌داده‌ها گزارش می‌شود. نتایج نشان می‌دهد که در اکثر مجموعه‌داده‌ها، مدل پیشنهادی بهترین نتایج را کسب کرده است. علاوه بر این، بهترین امتیاز Micro-F1 توسط MLFS-GLOCAL بدست آمد که در مجموعه‌داده‌های Arts، Corel5k، Recreation، Health، Entertainment و Social برتری قابل توجهی نسبت به روش دوم برتر داشت است. جداول نشان می‌دهند که مدل پیشنهادی در ۶۷ مورد از ۷۲ مورد مقایسه، اول و در بقیه موارد در رتبه‌دوم قرار دارد. نتایج تأیید می‌کنند که روش ما می‌تواند برای طیف گسترده‌ای از مجموعه‌های داده، برخلاف روش‌های دیگر، اعمال شود. به‌طور متوسط، ما بهبودهای قابل توجهی را در مقادیر این معیارها مشاهده کردیم، با ۰.۰۳۶۷ برای Macro-F1، ۰.۰۱۵۷ برای Micro-F1، ۰.۰۰۳۶۷ برای Coverage Error، ۰.۰۰۱۴ برای Hamming Loss و ۰.۰۰۱۱ برای F1 Ranking Loss.

علاوه‌براین، ما از نمودارهای رادری برای نشان دادن جامع بودن روش خود بر روی شش معیار ارزیابی مختلف در مقایسه با روش‌های دیگر استفاده می‌کنیم. این معیارها جنبه‌های مختلف کیفیت و اثربخشی روش‌ها را می‌سنجند [۴۲]. شایان ذکر است که معیارهای CVE، RPK و HML به منظور حفظ سازگاری با سایر معیارها به‌طور معکوس

جدول ۳.۴: ارزیابی مدل بر اساس معیار Macro-F1

مجموعه‌داده	SSFS	SCMFS	MIFS	CMFS	MRDM	NMDG	SCLS	LRFS	MDMR	MLFS-GLOCAL
Arts	0.102	0.1299	0.0760	0.0790	0.1301	0.0702	0.0532	0.0203	0.0638	0.1597
Business	0.0789	0.1066	0.0984	0.0762	0.0951	0.0633	0.0679	0.0427	0.0527	0.1224
Computers	0.1184	0.1205	0.0747	0.1128	0.1401	0.0514	0.0911	0.0503	0.0830	0.1609
Corel5k	0.0022	0.0032	0.0027	0.0024	0.0023	0.0019	0.0035	0.0018	0.0028	0.0044
Education	0.0936	0.1151	0.0607	0.0970	0.0880	0.0365	0.0488	0.0285	0.0451	0.1228
Entertainment	0.1736	0.1918	0.1291	0.1488	0.1907	0.1170	0.1191	0.0399	0.1138	0.2105
Health	0.1908	0.1991	0.1599	0.1844	0.1967	0.1405	0.1321	0.0673	0.1030	0.2283
Recreation	0.1317	0.1576	0.1347	0.1087	0.1691	0.1188	0.0802	0.0490	0.0905	0.1756
Reference	0.0940	0.1136	0.0892	0.0935	0.1093	0.0663	0.0690	0.0222	0.0613	0.1203
Science	0.0864	0.0977	0.0660	0.0784	0.0915	0.0538	0.0531	0.0310	0.0495	0.1090
Social	0.1349	0.1382	0.0974	0.1195	0.1177	0.0558	0.0537	0.0205	0.0611	0.1597
Society	0.1022	0.1114	0.0718	0.0693	0.1050	0.0750	0.0410	0.0262	0.0432	0.1190

جدول ۴.۴: ارزیابی مدل بر اساس معیار Average Precision

مجموعه‌داده	SSFS	SCMFS	MIFS	CMFS	MRDM	NMDG	SCLS	LRFS	MDMR	MLFS-GLOCAL
Arts	0.0988	0.0677	0.0670	0.0690	0.0694	0.0551	0.0588	0.0589	0.0581	0.0707
Business	0.0627	0.0630	0.0605	0.0623	0.0626	0.0584	0.0580	0.0553	0.0581	0.0689
Computers	0.0790	0.0688	0.0589	0.0717	0.0722	0.0551	0.0501	0.0524	0.0598	0.0807
Corel5k	0.0104	0.0103	0.0111	0.0104	0.0104	0.0107	0.0115	0.0100	0.0104	0.0118
Education	0.0581	0.0584	0.0561	0.0555	0.0640	0.0493	0.0532	0.0482	0.0493	0.0624
Entertainment	0.1056	0.1080	0.1039	0.1064	0.1085	0.1168	0.1124	0.0693	0.1041	0.1150
Health	0.0805	0.0819	0.0779	0.0860	0.0877	0.0724	0.0669	0.0715	0.0763	0.1105
Recreation	0.0788	0.0837	0.0837	0.0739	0.0898	0.0901	0.0690	0.0678	0.0708	0.1028
Reference	0.0486	0.0496	0.0414	0.0485	0.0486	0.0457	0.0429	0.0388	0.0433	0.0501
Science	0.0488	0.0490	0.0456	0.0515	0.0530	0.0399	0.0408	0.0403	0.0427	0.0535
Social	0.0608	0.0634	0.0503	0.0577	0.0613	0.0460	0.0494	0.0304	0.0497	0.0679
Society	0.0680	0.0703	0.1190	0.0656	0.0686	0.0697	0.0603	0.0604	0.0653	0.0709

جدول ۵.۴: ارزیابی مدل بر اساس معیار Ranking Loss

مجموعه‌داده	SSFS	SCMFS	MIFS	CMFS	MRDM	NMDG	SCLS	LRFS	MDMR	MLFS-GLOCAL
Arts	0.2066	0.2077	0.2090	0.2056	0.2012	0.2190	0.1998	0.2117	0.2124	0.1984
Business	0.0528	0.0488	0.0522	0.0526	0.0494	0.0547	0.0580	0.0583	0.0592	0.0490
Computers	0.1207	0.1172	0.1218	0.1208	0.1194	0.1204	0.1250	0.1222	0.1226	0.1148
Corel5k	0.2080	0.2078	0.2171	0.2080	0.2088	0.2162	0.2172	0.2146	0.2161	0.2063
Education	0.1211	0.1216	0.1274	0.1201	0.1231	0.1356	0.1353	0.1313	0.1374	0.1168
Entertainment	0.1664	0.1609	0.1690	0.1685	0.1641	0.1697	0.1801	0.1874	0.1726	0.1577
Health	0.0804	0.0807	0.0836	0.0803	0.0790	0.0890	0.0902	0.0979	0.1084	0.0789
Recreation	0.2209	0.2220	0.2201	0.2265	0.2297	0.2521	0.2693	0.2605	0.2575	0.2288
Reference	0.1009	0.0986	0.1043	0.1009	0.0982	0.1074	0.1116	0.1203	0.1176	0.0970
Science	0.1635	0.1583	0.1672	0.1639	0.1586	0.1739	0.2030	0.2015	0.2016	0.1559
Social	0.0707	0.0732	0.0825	0.0774	0.0741	0.0878	0.0863	0.1053	0.0957	0.0728
Society	0.1800	0.1813	0.1852	0.1865	0.1811	0.1902	0.1704	0.2058	0.2226	0.1758

جدول ٦.٤: ارزیابی مدل بر اساس معیار Hamming Loss

مجموعه داده	SSFS	SCMFS	MIFS	CMFS	MRDM	NMDG	SCLS	LRFS	MDMR	MLFS-GLOCAL
Arts	۰.۰۶۱۵	<u>۰.۰۵۸۸</u>	۰.۰۶۱۶	۰.۰۶۲۵	۰.۰۵۹۷	۰.۰۶۲۲	۰.۰۶۳۱	۰.۰۶۳۳	۰.۰۶۲۷	۰.۰۵۶۸
Business	۰.۰۲۸۸	<u>۰.۰۲۷۸</u>	۰.۰۲۸۱	۰.۰۲۹۰	۰.۰۲۸۶	۰.۰۲۸۷	۰.۰۲۸۶	۰.۰۲۸۸	۰.۰۲۸۸	۰.۰۲۷۶
Computers	<u>۰.۰۳۹۳</u>	۰.۰۳۹۴	۰.۰۴۰۴	۰.۰۴۰۲	<u>۰.۳۹۷۹</u>	۰.۰۴۱۱	۰.۰۳۹۶	۰.۰۴۲۲	۰.۰۴۱۲	۰.۰۳۸۸
Corel5k	۰.۰۰۹۵۱۷۰	۰.۰۰۹۵۰۶	<u>۰.۰۰۹۴۹۵</u>	۰.۰۰۹۵۰۶	۰.۰۰۹۵۰۴	۰.۰۰۹۵۰۶	۰.۰۰۹۵۱۶	۰.۰۰۹۵۱۲	۰.۰۰۹۵۳۵	۰.۰۰۹۴۹۵
Education	۰.۰۴۱۴	<u>۰.۰۴۱۰</u>	۰.۰۴۲۷	۰.۰۴۱۸	۰.۰۴۳۶	۰.۰۴۰۵	۰.۰۴۲۳	۰.۰۴۴۴	۰.۰۴۴۲	۰.۰۴۹۲
Entertainment	۰.۰۶۱۵	<u>۰.۰۶۹۴</u>	۰.۰۶۴۱	۰.۰۶۳۰	۰.۰۶۱۳	۰.۰۶۰۸	۰.۰۶۳۴	۰.۰۶۷۶	۰.۰۶۲۹	۰.۰۵۰۵
Health	۰.۰۴۲۷	<u>۰.۰۴۰۷</u>	۰.۰۴۳۶	۰.۰۴۳۳	۰.۰۴۲۳	۰.۰۴۳۰	۰.۰۴۶۱	۰.۰۵۰۶	۰.۰۴۶۶	۰.۰۳۸۲
Recreation	۰.۰۶۱۷	<u>۰.۰۵۹۲</u>	۰.۰۶۲۰	۰.۰۶۲۳	<u>۰.۰۵۹۰</u>	۰.۰۶۱۰	۰.۰۶۲۴	۰.۰۶۴۷	۰.۰۶۲۸	۰.۰۵۷۱
Reference	۰.۰۲۹۶	<u>۰.۰۲۸۷</u>	۰.۰۲۹۷	۰.۰۲۹۷	۰.۰۲۹۱	۰.۰۲۹۴	۰.۰۳۲۳	۰.۰۳۴۷	۰.۰۳۲۲	۰.۰۲۷۵
Science	۰.۰۳۴۹	<u>۰.۰۳۴۲</u>	۰.۰۳۵۸	۰.۰۳۵۱	۰.۰۳۴۶	۰.۰۳۵۲	۰.۰۳۵۷	۰.۰۳۵۸	۰.۰۳۵۴	۰.۰۳۳۶
Social	۰.۰۲۴۳	<u>۰.۰۲۳۶</u>	۰.۰۲۵۳	۰.۰۲۴۷	۰.۰۲۴۲	۰.۰۲۸۱	۰.۰۲۷۵	۰.۰۳۱۳	۰.۰۲۶۲	۰.۰۲۲۱
Society	۰.۰۵۶۳	<u>۰.۰۵۵۳</u>	۰.۰۵۷۱	۰.۰۵۷۴	۰.۰۵۰۴	۰.۰۵۷۶	۰.۰۵۹۰	۰.۰۵۸۳	۰.۰۵۴۰	

جدول ٧.٤: ارزیابی مدل بر اساس معیار Coverage Error

مجموعه داده	SSFS	SCMFS	MIFS	CMFS	MRDM	NMDG	SCLS	LRFS	MDMR	MLFS-GLOCAL
Arts	۷.۸۲۰	۷.۸۵۲	۷.۹۳۰	۷.۸۲۰	۷.۷۱۰	۸.۲۱۷	<u>۷.۶۶۲</u>	۸.۰۴۵	۸.۰۵۹	۷.۶۲۷
Business	۲.۷۷۱	<u>۲.۵۸۴</u>	۲.۶۸۶	۲.۷۴۰	۲.۶۰۹	۲.۸۰۸	۲.۹۳۳	۴.۰۰۹	۴.۰۶۰	<u>۲.۵۸۸</u>
Computers	۶.۴۴۰	۶.۴۲۰	۶.۴۵۸	<u>۶.۴۹۵</u>	<u>۶.۲۲۴</u>	۶.۶۱۹	۶.۷۲۳	۶.۸۴۹	۶.۹۷۳	۶.۲۳۷
Corel5k	۱۶۴.۱۹	<u>۳۶.۱۶۴</u>	۱۷۰.۹۰	۱۶۴.۸۳	۱۶۴.۸۱	۱۷۲.۰۲	۱۷۱.۹۵	۱۷۱.۹۶	۱۷۲.۰۰	۱۶۳.۶۳
Education	۰.۸۷۰	۰.۸۸۲	۶.۰۹۳	<u>۰.۸۲۰</u>	۶.۰۱۹	۶.۴۵۳	۶.۵۰۵	۶.۴۱۲	۶.۶۲۱	۰.۷۳۵
Entertainment	۰.۱۹۰	۰.۰۷۵	۰.۲۶۹	۰.۲۰۴	۰.۱۹۹	<u>۰.۰۲۷</u>	۰.۶۱۰	۰.۶۹۴	۰.۳۵۹	۰.۰۱۰
Health	۴.۹۰۰	۴.۹۰۵	۴.۰۹۶	۴.۹۰۰	<u>۴.۹۳۲</u>	۴.۹۳۲	۴.۳۷۸	۴.۶۶۳	۴.۳۴۲	۴.۹۰۳
Recreation	۷.۱۲۵	۶.۹۴۴	۷.۲۳۱	۷.۲۰۹	<u>۶.۸۸۲</u>	۷.۴۰۶	۷.۱۲۴	۷.۹۷۸	۷.۰۱۱	۷.۸۳۰
Reference	۴.۸۲۰	۴.۷۴۲	۴.۹۴۶	۴.۸۱۱	<u>۴.۷۷۴</u>	۵.۰۲۹	۰.۱۹۴	۰.۴۷۰	۰.۴۱۱	۴.۶۹۸
Science	۸.۸۶۷	۸.۶۳۰	۹.۰۴۰	۸.۸۸۵	<u>۸.۶۱۳</u>	۹.۴۳۰	۱۰.۶۹۸	۱۰.۶۳۷	۱۰.۶۲۴	۸.۵۲۵
Social	۴.۸۲۰	<u>۴.۷۳۹</u>	۵.۱۳۱	۴.۹۱۱	۴.۷۶۱	۵.۳۴۱	۵.۳۳۹	۶.۱۱۶	۵.۸۴۱	۴.۷۱۹
Society	۷.۶۲۱	۷.۶۴۰	۷.۷۵۴	۷.۷۷۰	۷.۶۱۳	۷.۱۹۹	<u>۷.۰۷۷</u>	۸.۰۳۶	۷.۰۸۰	۷.۴۷۰

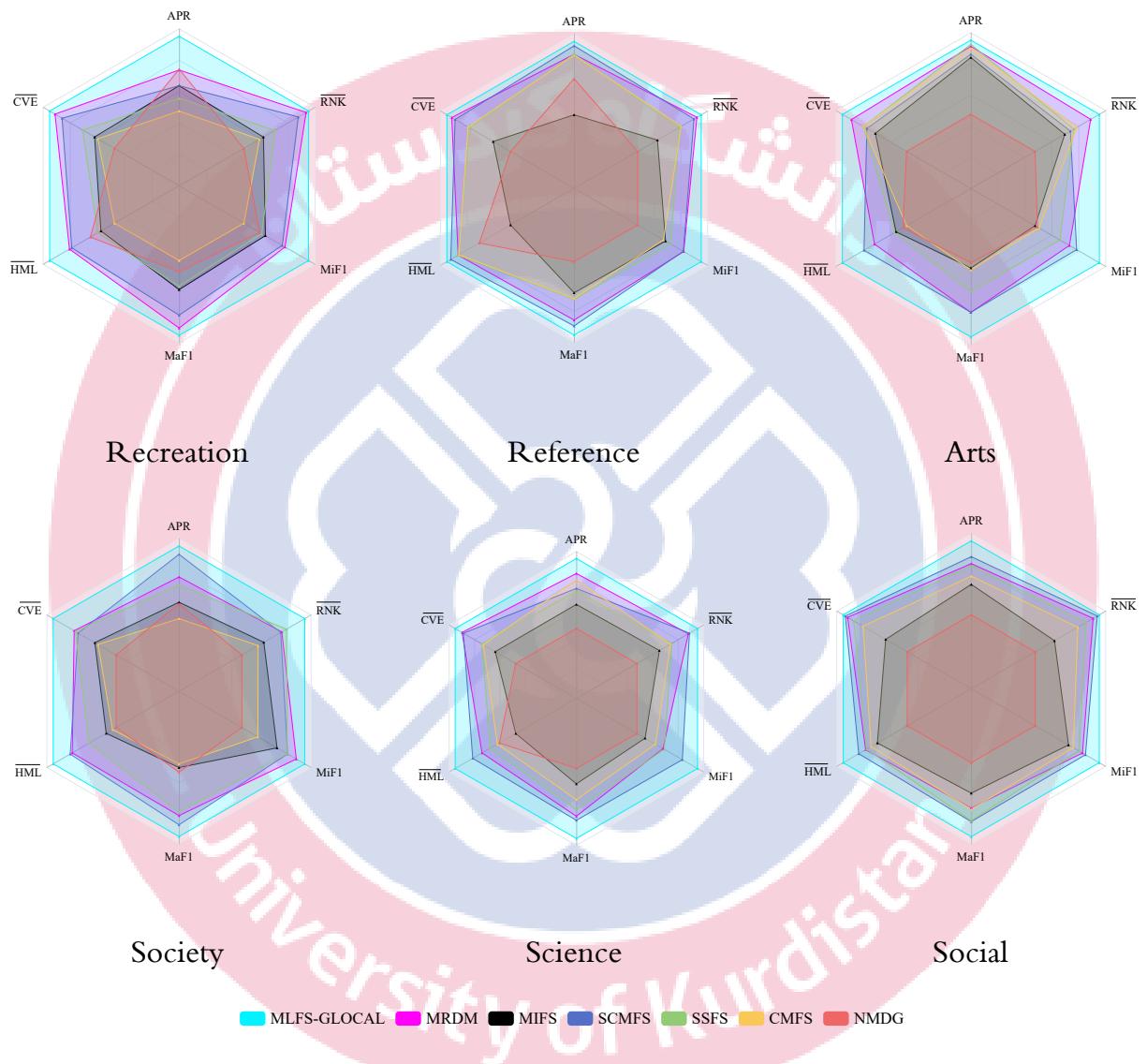
مورد استفاده قرار می‌گیرند، به طوری که مقدار بزرگتر نشان دهنده عملکرد بهتر می‌باشد. همچنین برای اینکه مقایسه منصفانه و واضح باشد، داده‌های شکل ۱.۴ را نرمال می‌کنیم، به طوری که همه مقادیر بین ۰ و ۵٪ باشد. هر چه یک روش مساحت بیشتری را در نمودارهای رadar پوشش دهد، به این معنی است که در تمام معیارهای ارزیابی بهتر است. شکل ۱.۴ نشان می‌دهد که روش ما نسبت به روش‌های دیگر مساحت بیشتری دارد، به این معنی که از نظر عملکرد و معیارهای ارزیابی جامع‌تر و برتر است، این نشان می‌دهد که روش ما می‌تواند انواع مختلف مشکلات و موقعیت‌ها را بهتر از روش‌های موجود مدیریت کند. عملکرد MLFS-GLOCAL و سایر رویکردهای مقایسه‌ای به صورت بصری با استفاده از چهار مجموعه داده نشان داده شده است: Arts, Corel5k, Business و Entertainment. محور x در شکل‌های ۲.۴ و ۵.۴ عملکرد معیارهای ارزیابی مختلف را نشان می‌دهد، در حالی که محور x درصد ویژگی‌های انتخاب شده را نشان می‌دهد. روش ما در تعداد کم ویژگی بهتر است و در شکل‌های ۲.۴ و ۵.۴، واضح است که هر روش با انتخاب ویژگی‌های بیشتر بهتر عمل می‌کند. علاوه بر این، ما می‌توانیم عملکرد هر روش را روی همان مجموعه داده با همان معیار مقایسه کنیم. به عنوان مثال، در شکل ۲.۴ نتایج را برای مجموعه داده‌های Arts, Corel5k, Business و Entertainment نشان می‌دهد، که در آن معیارهای الگوریتم MLFS-GLOCAL به طور قابل توجهی بهتر از چند الگوریتم دیگر است.

۵.۴ تجزیه و تحلیل پارامترها

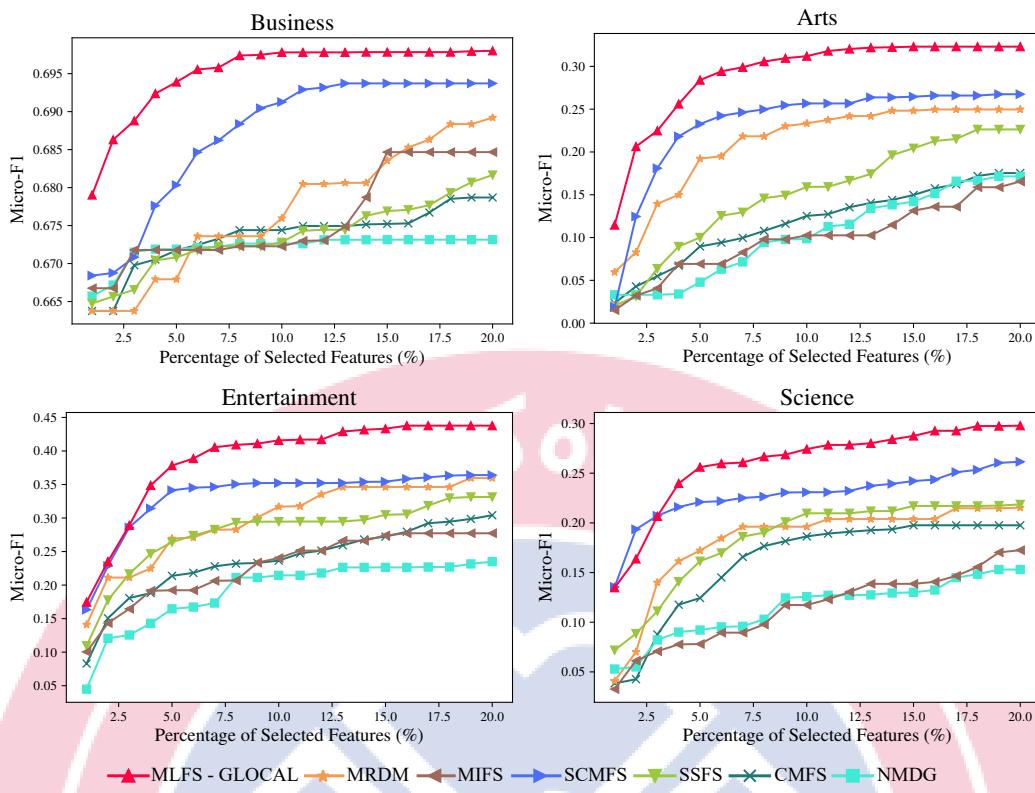
در این بخش، تأثیر پارامترها، از جمله پارامتر منظم‌ساز گراف λ_1 ، پارامتر همبستگی برچسب سراسری و محلی λ_2 ، و پارامتر خلوتی λ_3 را تحلیل می‌کنیم. شکل‌های ۶.۴ و ۷.۴ Macro-F1، Micro-F1 و Hamming Loss و Ranking Loss روش ما را با شش مجموعه داده λ_1 ، λ_2 و λ_3 نشان می‌دهند. این شکل‌ها به صورت سه‌بعدی ترسیم شده‌اند، یعنی سه محور مربوط به λ_1 ، λ_2 و λ_3 است [۴۳].

• پارامتر λ_1

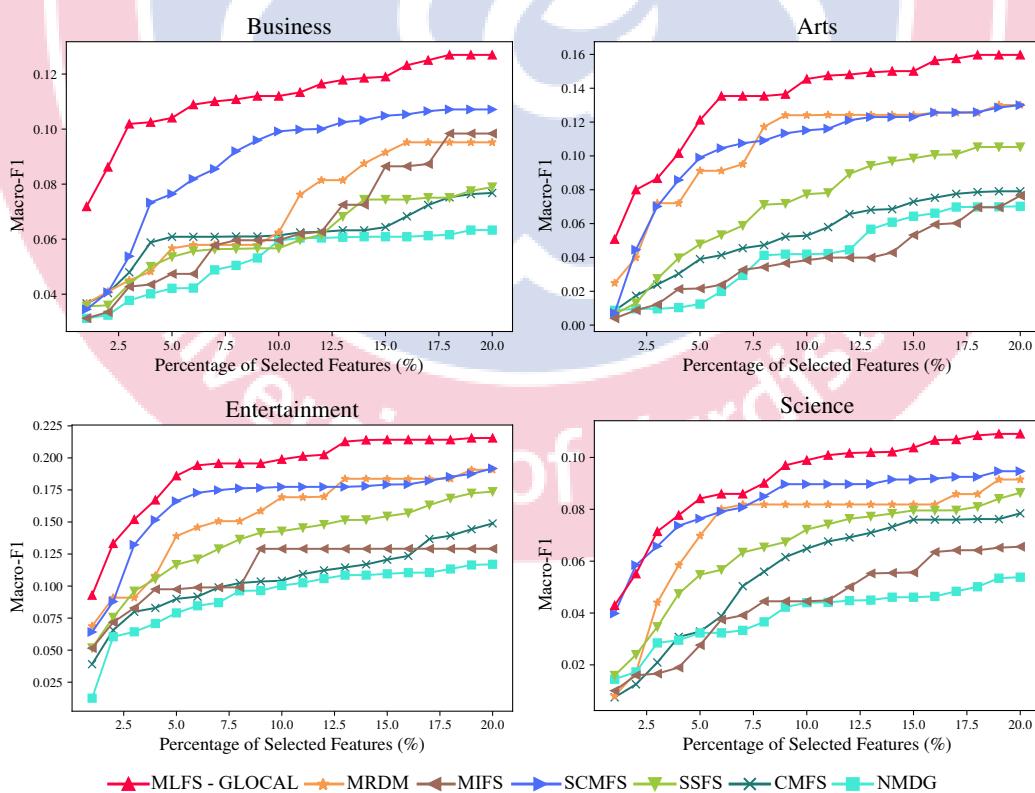
این پارامتر اثربخشی منظم‌ساز خمینه فضای نمونه در مدل را کنترل می‌کند. در این تحلیل پارامتر، مقادیر موجود در مجموعه $\{0, 10, 100, 1/10, 1/100\}$ برای پارامتر λ_1 در همه مجموعه‌های داده انتخاب شده‌اند. از شکل‌های ۶.۴ و ۷.۴ می‌توان نتیجه گرفت که مقدار بهینه برای این پارامتر معمولاً کمتر از ۱ است.



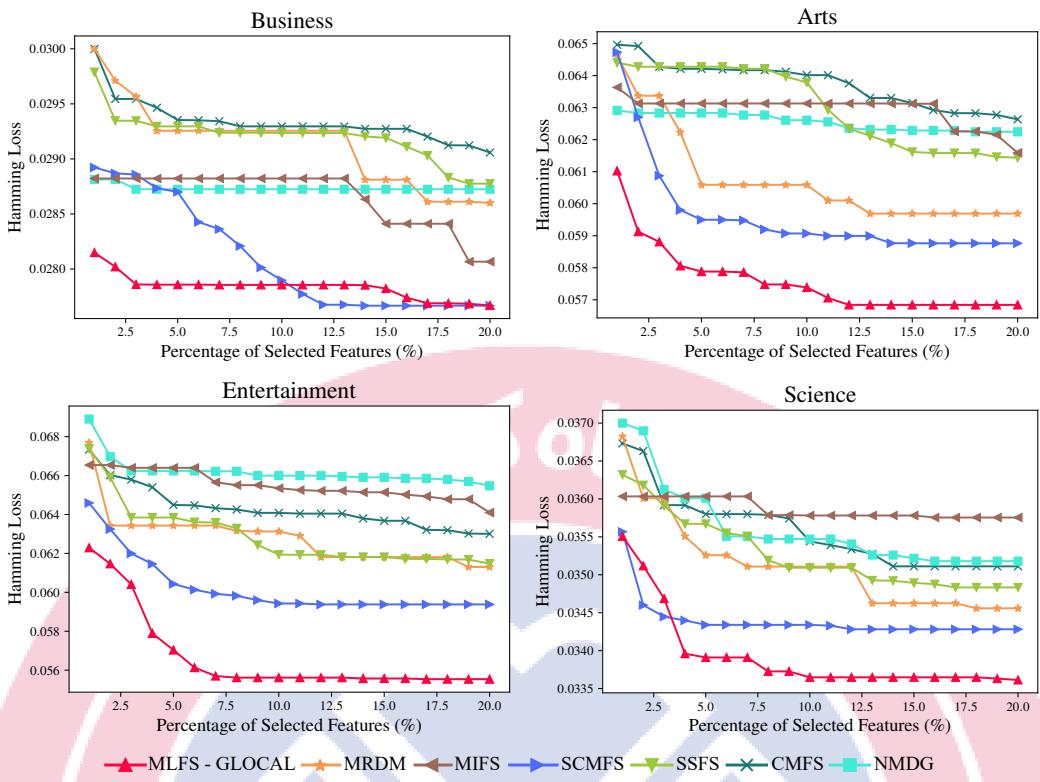
شكل ١٠.٤: نمودار رادار در شش معیار مختلف



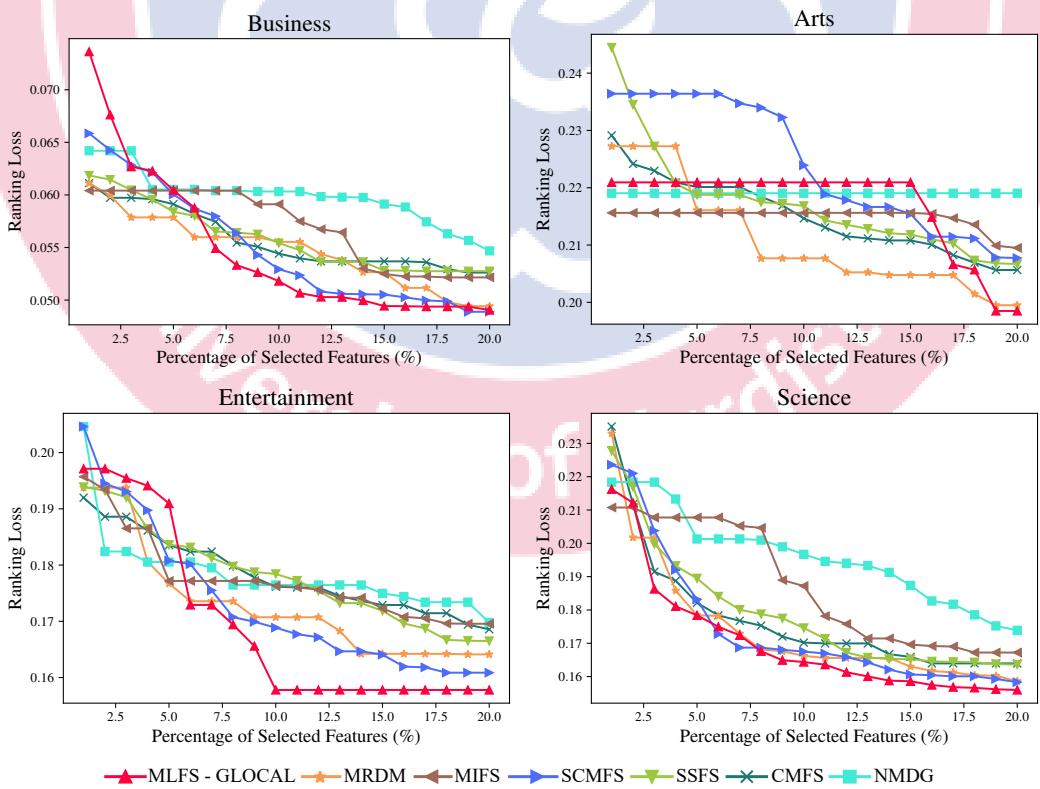
شکل ۲۰.۴: نمودار درصد انتخاب ویژگی‌ها در معیار Micro-F1



شکل ۲۰.۴: نمودار درصد انتخاب ویژگی‌ها در معیار Macro-F1



شکل ۴.۴: نمودار درصد انتخاب ویژگی‌ها در معیار Hamming Loss



شکل ۵.۴: نمودار درصد انتخاب ویژگی‌ها در معیار Ranking Loss

• پارامتر₂

این پارامتر اثربخشی همبستگی بر حسب سراسری و محلی مدل را نشان می‌دهد و مقادیر تحلیل شده برای پارامتر λ_2 { 10^{-6} ، 10^{-12} ، $0.1/0$ ، $0.01/0$ } هستند. همان‌طور که در شکل‌های ۶.۴ و ۷.۴ مشاهده می‌کنیم، λ_2 با مقادیر کوچک معمولاً عملکرد بهتری را از نظر Macro-F1 و Micro-F1 نشان می‌دهد و با مقادیر بالا معمولاً از نظر معیارهای Coverage Error و Hamming Loss در چهار مجموعه‌داده عملکرد بهتری دارد. مقادیر نزدیک به صفر یا مقادیر بسیار بزرگ برای این پارامتر ممکن است عملکرد نسبتاً خوبی نداشته باشند.

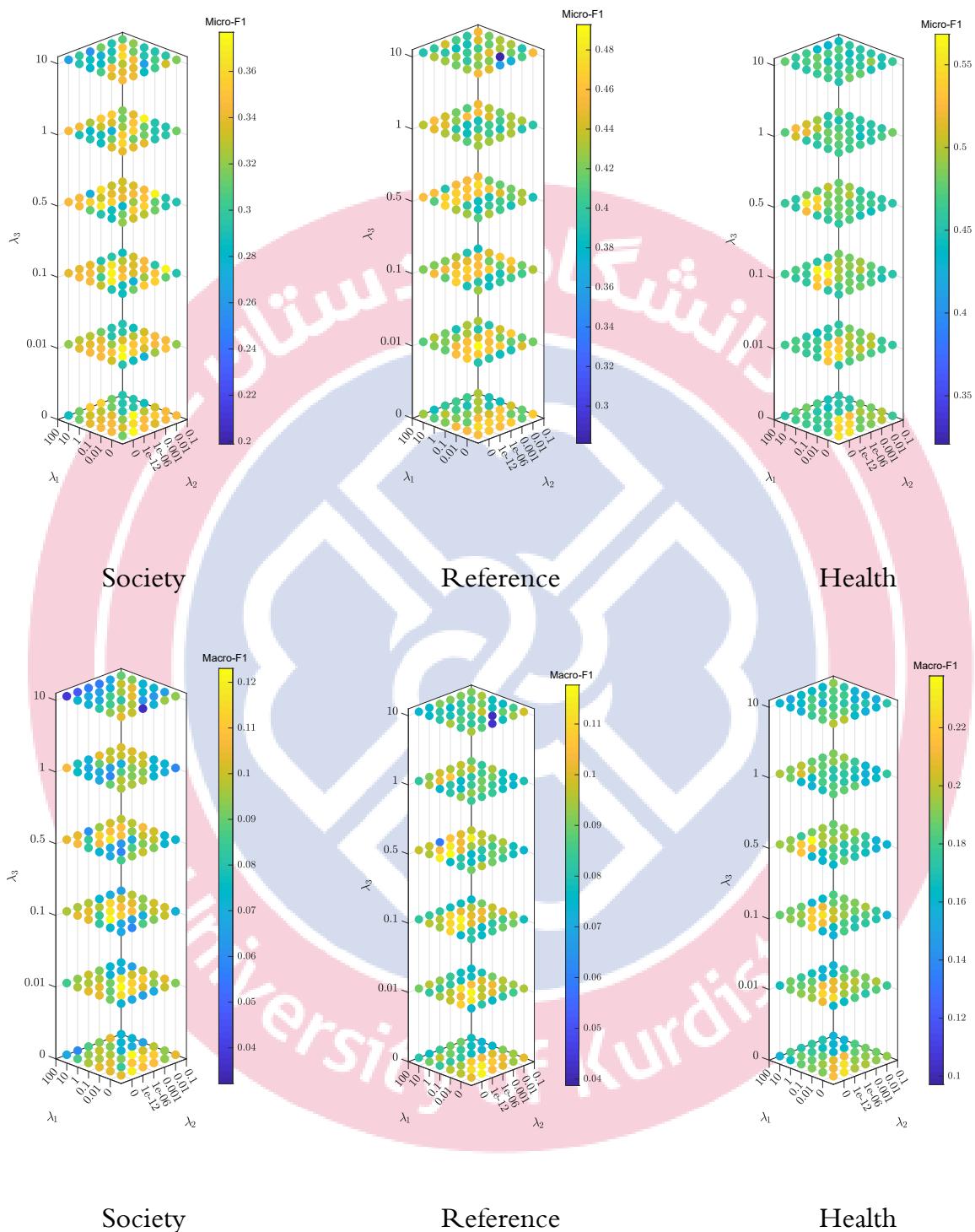
• پارامتر₃

عبارت منظم‌ساز خلوت در MLFS-GLOCAL توسط λ_3 کنترل می‌شود. برای پارامتر λ_3 محدوده مقادیر { 10^{-5} ، $0.1/0$ ، $0.01/0$ } است. نتایج نشان می‌دهد که λ_3 یک کمیت ظرفی است که معمولاً نیاز به تنظیم دقیق دارد. نتایج نشان می‌دهد که انتخاب مقادیر زیر ۱ برای این پارامتر بهتر است.

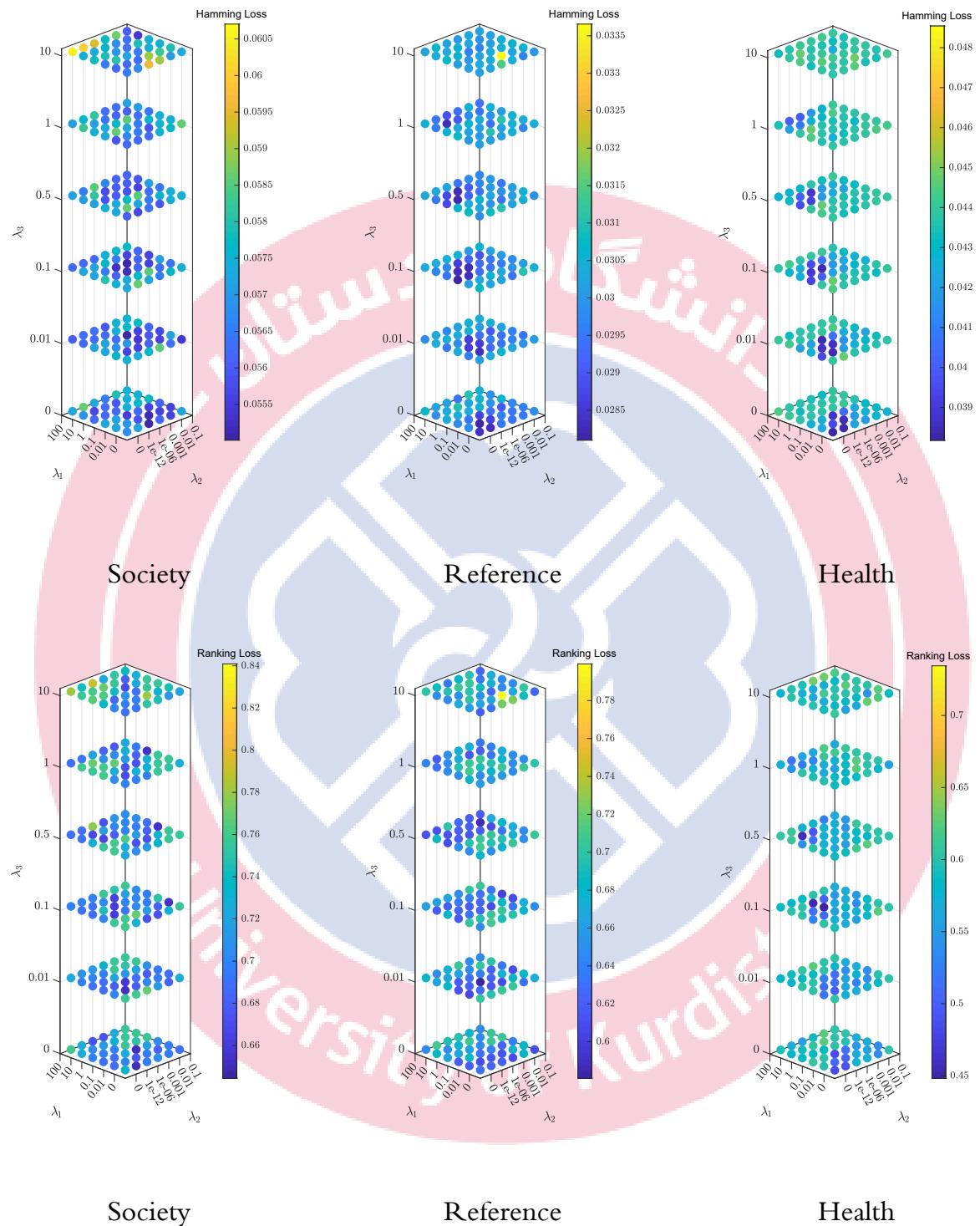
۶.۴ تحلیل همگرایی

در این بخش، ما رفتار همگرایی^۱ مدل پیشنهادی (۱۰.۳) را با انجام آزمایش‌هایی بر روی چهار مجموعه‌داده Arts، Reference و Health ارزیابی می‌کنیم. الگوریتم ۱ را برای هر مجموعه‌داده با ۶۰۰ تکرار اجرا می‌کنیم. مقدار تابع هدف را در برابر تعداد تکرارها در شکل ۸.۴ رسم کرده‌ایم تا نشان دهیم چگونه روش ما همگرا می‌شود. همان‌طور که از شکل ۸.۴ مشاهده می‌شود، مقدار تابع هدف به سرعت و به طور پیوسته در تکرارهای اولیه کاهش می‌یابد، که نشان می‌دهد روش ما به سرعت به یک راه حل بهینه نزدیک می‌شود. در تکرارهای بعدی، مقدار تابع هدف بسیار کم تغییر می‌کند، و این نشان می‌دهد روش ما به یک راه حل تقریباً بهینه رسیده است. این بیانگر این می‌باشد که روش ارائه شده عملکرد همگرایی سریع و پایداری دارد.

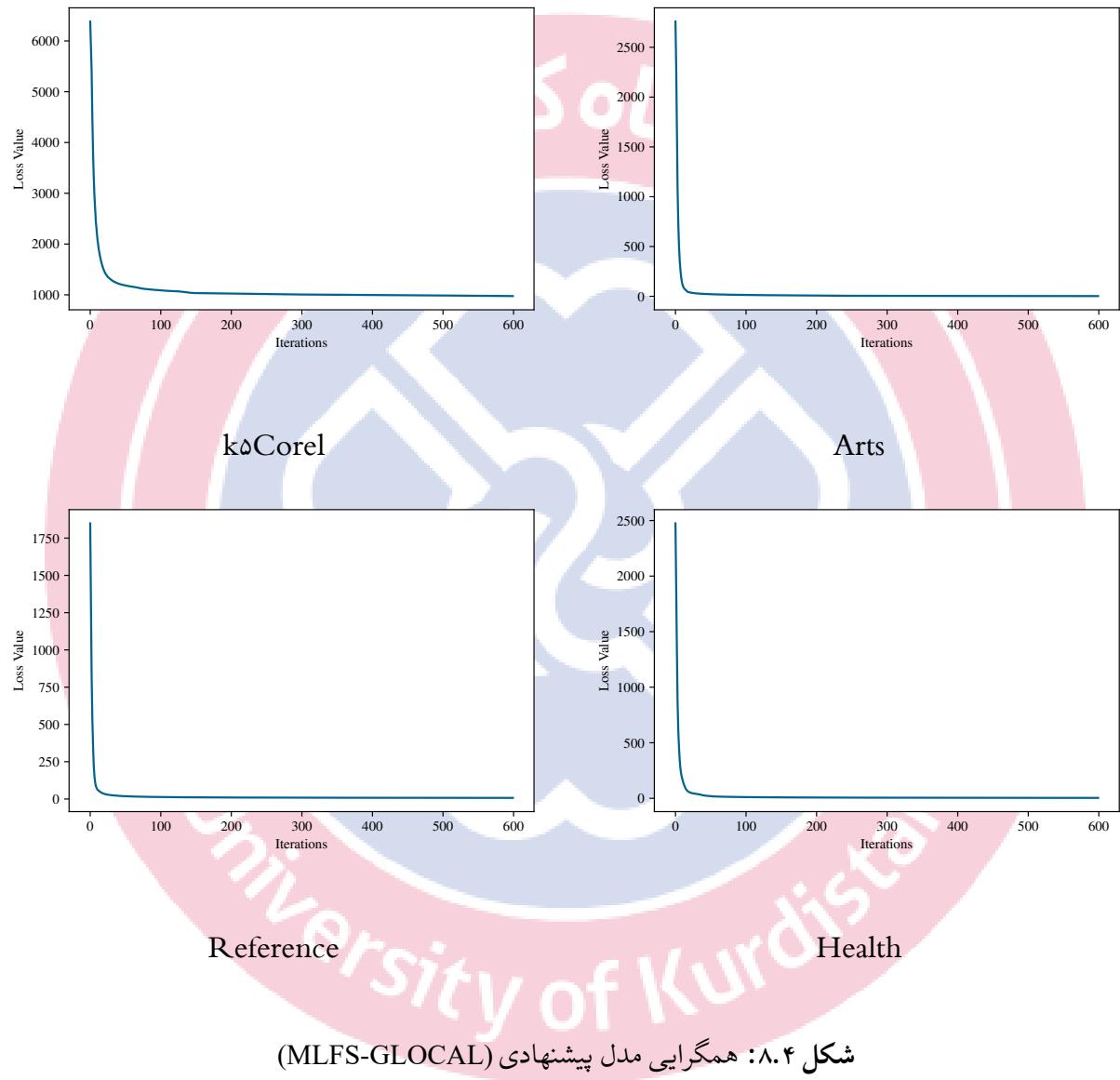
¹ Convergence



شکل ۶.۴: تجزیه و تحلیل پارامترهای α و β در مدل پیشنهادی



شکل ۷.۴: تجزیه و تحلیل پارامترهای α و β در مدل پیشنهادی



شکل ۸.۴: همگرایی مدل پیشنهادی (MLFS-GLOCAL)



فصل پنجم

نتیجه گیری و کارهای آینده

۱۰۵ نتیجه‌گیری

در این پایان‌نامه، ما یک روش جدید انتخاب ویژگی چندبرچسبه را با درنظر گرفتن همبستگی برچسب سراسری و محلی پیشنهاد کردیم. همبستگی برچسب سراسری به بهره‌برداری از ساختار زیربنایی فضای برچسب کمک می‌کند. این به مدل اجازه می‌دهد تا روابط و همبستگی‌های بین برچسبها را یاد بگیرید، از سوی دیگر، همبستگی محلی برچسب به ارتباط بین برچسبها در یک پارتویشن محلی خاص اشاره دارد. با درنظر گرفتن همبستگی بین برچسبها و گنجاندن آن‌ها در بازنمایی ویژگی، مدل ویژگی‌هایی را شناسایی می‌کند که بیشترین ارتباط را با برچسبها دارند. علاوه‌براین، MLFS-GLOCAL فضای مشترک بین ماتریس ویژگی و ماتریس برچسب را می‌آموزد تا همبستگی برچسب ضمنی را استخراج کند و انتخاب ویژگی را راهنمایی کند. این ماتریس کم‌بعد از طریق منظم‌ساز چندگانه محدود می‌شود، به این معنی که ساختار محلی مشابهی با داده‌های اصلی دارد و اطلاعات ارزشمند در داده‌ها حفظ می‌شود. یک الگوریتم تکراری مبتنی بر بهینه‌سازی متناوب برای حل تابع هدف با منظم‌ساز $\ell_{2,1}$ ایجاد شده است. در نهایت، آزمایش‌های گسترشده بر روی مجموعه‌داده‌های چندبرچسبه انجام شده است تا اثربخشی مدل پیشنهادی را در برابر تعدادی از روش‌های انتخاب ویژگی پیشرفته نشان دهد.

۲۰۵ پیشنهادهایی برای تحقیقات آتی

روش‌های پیشنهادی آتی این پایان‌نامه می‌تواند در زمینه‌های زیر باشد:

- با توجه به اهمیت اطلاعات برچسب در دنیای واقعی می‌توان بر روی داده‌های نیمه‌ناظارتی^۱ [۴۴، ۴۵] یا گم‌شده [۴۶] براساس همبستگی بین برچسب‌ها تمرکز کرد.
- می‌توان این روش پیشنهادی را بر روی داده‌های چندناییه^۲ چندبرچسبه توسعه داد که در آن داده‌ها با مجموعه ویژگی‌ها یا نماهای متعدد نشان داده می‌شوند و هر نمونه با چندین برچسب مرتبط می‌باشد.

1. Semi-supervised

2. Multi-view



- [1] J. Liu, Y. Lin, Y. Li, W. Weng, and S. Wu, “Online multi-label streaming feature selection based on neighborhood rough set,” *Pattern Recognition*, vol. 84, pp. 273–287, 2018.
- [2] F. Li, D. Miao, and W. Pedrycz, “Granular multi-label feature selection based on mutual information,” *Pattern Recognition*, vol. 67, pp. 410–423, 2017.
- [3] Y. Lin, Q. Hu, J. Liu, J. Li, and X. Wu, “Streaming feature selection for multilabel learning based on fuzzy mutual information,” *IEEE Transactions on Fuzzy Systems*, vol. 25, no. 6, pp. 1491–1507, 2017.
- [4] W. Gao, L. Hu, and P. Zhang, “Class-specific mutual information variation for feature selection,” *Pattern Recognition*, vol. 79, pp. 328–339, 2018.
- [5] J. Huang, G. Li, Q. Huang, and X. Wu, “Joint feature selection and classification for multilabel learning,” *IEEE transactions on cybernetics*, vol. 48, no. 3, pp. 876–889, 2017.
- [6] P. Zhu, Q. Xu, Q. Hu, C. Zhang, and H. Zhao, “Multi-label feature selection with missing labels,” *Pattern Recognition*, vol. 74, pp. 488–502, 2018.
- [7] R. Hong, M. Wang, Y. Gao, D. Tao, X. Li, and X. Wu, “Image annotation by multiple-instance learning with discriminative feature mapping and selection,” *IEEE transactions on cybernetics*, vol. 44, no. 5, pp. 669–680, 2013.
- [8] J. Xie, M. Wang, S. Xu, Z. Huang, and P. W. Grant, “The unsupervised feature selection algorithms based on standard deviation and cosine similarity for genomic data analysis,” *Frontiers in Genetics*, vol. 12, p. 684100, 2021.
- [9] P. Wang and C. Domeniconi, “Building semantic kernels for text classification using wikipedia,” in *ACM SIGKDD international conference on Knowledge discovery and data mining*, 2008, pp. 713–721.
- [10] K. Trohidis, G. Tsoumakas, G. Kalliris, and I. Vlahavas, “Multi-label classification of music by emotion,” *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 2011, no. 1, pp. 1–9, 2011.
- [11] M.-L. Zhang and Z.-H. Zhou, “A review on multi-label learning algorithms,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, no. 8, pp. 1819–1837, 2014.
- [12] Y. Zhang and Y. Ma, “Non-negative multi-label feature selection with dynamic graph constraints,” *Knowledge-Based Systems*, vol. 238, p. 107924, 2022.

- [13] L. Jian, J. Li, K. Shu, and H. Liu, “Multi-label informed feature selection,” in *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*, 2016, pp. 1627–1633.
- [14] Z.-H. Zhou, “A brief introduction to weakly supervised learning,” *National science review*, vol. 5, no. 1, pp. 44–53, 2018.
- [15] J. Fürnkranz, E. Hüllermeier, E. Loza Mencía, and K. Brinker, “Multilabel classification via calibrated label ranking,” *Machine learning*, vol. 73, pp. 133–153, 2008.
- [16] S. Ji, L. Tang, S. Yu, and J. Ye, “Extracting shared subspace for multi-label classification,” in *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2008, pp. 381–389.
- [17] J. Read, B. Pfahringer, G. Holmes, and E. Frank, “Classifier chains for multi-label classification,” *Machine learning*, vol. 85, pp. 333–359, 2011.
- [18] S.-J. Huang and Z.-H. Zhou, “Multi-label learning by exploiting label correlations locally,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 26, no. 1, 2012, pp. 949–955.
- [19] W. Weng, Y. Lin, S. Wu, Y. Li, and Y. Kang, “Multi-label learning based on label-specific features and local pairwise label correlation,” *Neurocomputing*, vol. 273, pp. 385–394, 2018.
- [20] D. D. Lee and H. S. Seung, “Learning the parts of objects by non-negative matrix factorization,” *Nature*, vol. 401, no. 6755, pp. 788–791, 1999.
- [21] Y. Lin, Q. Hu, J. Liu, and J. Duan, “Multi-label feature selection based on max-dependency and min-redundancy,” *Neurocomputing*, vol. 168, pp. 92–103, 2015.
- [22] J. Lee and D.-W. Kim, “Scls: Multi-label feature selection based on scalable criterion for large label set,” *Pattern Recognition*, vol. 66, pp. 342–352, 2017.
- [23] P. Zhang, G. Liu, and W. Gao, “Distinguishing two types of labels for multi-label feature selection,” *Pattern Recognition*, vol. 95, pp. 72–82, 2019.
- [24] F. Nie, H. Huang, X. Cai, and C. Ding, “Efficient and robust feature selection via joint l_{2,1}-norms minimization,” in *Advances in Neural Information Processing Systems*, vol. 23, 2010, pp. 1813–1821.
- [25] X. Cai, F. Nie, and H. Huang, “Exact top-k feature selection via l_{2,0}-norm constraint,” in *PIInternational Joint Conference on Artificial Intelligence*, 2013, p. 1240–1246.

- [26] A. Braytee, W. Liu, D. R. Catchpoole, and P. J. Kennedy, “Multi-label feature selection using correlation information,” in *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, 2017, pp. 1649–1656.
- [27] L. Hu, Y. Li, W. Gao, P. Zhang, and J. Hu, “Multi-label feature selection with shared common mode,” *Pattern Recognition*, vol. 104, p. 107344, 2020.
- [28] W. Gao, Y. Li, and L. Hu, “Multilabel feature selection with constrained latent structure shared term,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 34, no. 3, pp. 1253–1262, 2023.
- [29] R. Huang and Z. Wu, “Multi-label feature selection via manifold regularization and dependence maximization,” *Pattern Recognition*, vol. 120, p. 108149, 2021.
- [30] Y. Fan, J. Liu, P. Liu, Y. Du, W. Lan, and S. Wu, “Manifold learning with structured subspace for multi-label feature selection,” *Pattern Recognition*, vol. 120, p. 108169, 2021.
- [31] Y. Fan, J. Liu, W. Weng, B. Chen, Y. Chen, and S. Wu, “Multi-label feature selection with local discriminant model and label correlations,” *Neurocomputing*, vol. 442, pp. 98–115, 2021.
- [32] Y. Zhu, J. T. Kwok, and Z.-H. Zhou, “Multi-label learning with global and local label correlation,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 30, no. 6, pp. 1081–1094, 2017.
- [33] D. Zhao, Q. Gao, Y. Lu, and D. Sun, “Learning multi-label label-specific features via global and local label correlations,” *Soft Computing*, vol. 26, no. 5, pp. 2225–2239, 2022.
- [34] S. Kumar and R. Rastogi, “Low rank label subspace transformation for multi-label learning with missing labels,” *Information Sciences*, vol. 596, pp. 53–72, 2022.
- [35] 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, 2010.
- [36] H.-Y. Chuang, E. Lee, Y.-T. Liu, D. Lee, and T. Ideker, “Network-based classification of breast cancer metastasis,” *Molecular systems biology*, vol. 3, no. 1, p. 140, 2007.
- [37] A. Subramanian, P. Tamayo, V. K. Mootha, S. Mukherjee, B. L. Ebert, M. A. Gillette, A. Paulovich, S. L. Pomeroy, T. R. Golub, E. S. Lander *et al.*, “Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles,” *Proceedings of the National Academy of Sciences*, vol. 102, no. 43, pp. 15 545–15 550, 2005.

- [38] G. Doquire and M. Verleysen, “Mutual information-based feature selection for multilabel classification,” *Neurocomputing*, vol. 122, pp. 148–155, 2013.
- [39] M.-L. Zhang and Z.-H. Zhou, “Ml-knn: A lazy learning approach to multi-label learning,” *Pattern recognition*, vol. 40, no. 7, pp. 2038–2048, 2007.
- [40] J. Liu, Y. Lin, S. Wu, and C. Wang, “Online multi-label group feature selection,” *Knowledge-Based Systems*, vol. 143, pp. 42–57, 2018.
- [41] J. Zhang, Z. Luo, C. Li, C. Zhou, and S. Li, “Manifold regularized discriminative feature selection for multi-label learning,” *Pattern Recognition*, vol. 95, pp. 136–150, 2019.
- [42] S. A. Seyedi, S. S. Ghodsi, F. Akhlaghian, M. Jalili, and P. Moradi, “Self-paced multi-label learning with diversity,” in *Asian Conference on Machine Learning*. PMLR, 2019, pp. 790–805.
- [43] N. Salahian, F. A. Tab, S. A. Seyedi, and J. Chavoshinejad, “Deep autoencoder-like nmf with contrastive regularization and feature relationship preservation,” *Expert Systems with Applications*, vol. 214, p. 119051, 2023.
- [44] J. Chavoshinejad, S. A. Seyedi, F. Akhlaghian Tab, and N. Salahian, “Self-supervised semi-supervised nonnegative matrix factorization for data clustering,” *Pattern Recognition*, vol. 137, p. 109282, 2023.
- [45] S. A. Seyedi, P. Moradi, and F. A. Tab, “A weakly-supervised factorization method with dynamic graph embedding,” in *2017 Artificial Intelligence and Signal Processing Conference (AISP)*, 2017, pp. 213–218.
- [46] S. A. Seyedi, F. Akhlaghian Tab, A. Lotfi, N. Salahian, and J. Chavoshinejad, “Elastic adversarial deep nonnegative matrix factorization for matrix completion,” *Information Sciences*, vol. 621, pp. 562–579, 2023.

Abstract

In various application domains, high-dimensional multi-label data has become more prevalent, presenting two significant challenges: instances with high-dimensional features and a large number of labels. In the context of multi-label feature selection, the objective is to choose a subset of features from a given set that is highly pertinent for predicting multiple labels or categories associated with each instance. However, certain characteristics of multi-label classification, such as label dependencies and imbalanced label distribution, have often been overlooked although hold valuable insights for designing effective multi-label feature selection algorithms. In this research, we propose a feature selection model which exploits explicit global and local label correlations to select discriminative features across multiple labels. In addition, by representing the feature matrix and label matrix in a shared latent space, the model aims to capture the underlying correlations between features and labels. The shared representation can reveal common patterns or relationships that exist across multiple labels and features. An objective function involving $\ell_{2,1}$ -norm regularization is formulated, and an alternating optimization-based iterative algorithm is designed to obtain the sparse coefficients for multi-label feature selection. The proposed method was evaluated on twelve real-world multi-label datasets using six evaluation metrics, through comprehensive experiments. The results indicate its effectiveness, surpassing that of several representative methods.

Keywords: feature selection, multi-label learning, label correlation, nonnegative matrix factorization



University of Kurdistan

Faculty of Engineering

Department of Computer Software Engineering

A Thesis Submitted to the Postgraduate Studies Office in Partial Fulfillment of the Requirements
for the Degree of M.Sc. in Computer Engineering - Artificial Intelligence and Robotics

Title:

Multi-Label Feature Selection by Exploiting Global and Local Label Correlation

By:

Mohammad Faraji

The above thesis was evaluated and approved by the following members of the thesis committee –
Excellent quality on September 12, 2023.

<u>Position</u>	<u>Name</u>	<u>Academic Rank</u>	<u>Signature</u>
1. Supervisor:	Dr. Fardin Akhlaghian Tab	Associate Prof.	
2. Advisor :	Seyed Amjad Seyed		
2. External Examiner:	Dr. Mohsen Ramezani Tab	Assistant Prof.	
3. Internal Examiner:	Dr. Rojjar PirMohamadiani	Assistant Prof.	

Head of Department:

Faculty Graduate Coordinator:



University of Kurdistan

Faculty of Engineering

Department of Computer Software Engineering

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree
of M.Sc. in Computer Engineering – Artificial Intelligence and Robotics

Title:

**Multi-Label Feature Selection by Exploiting Global and Local
Label Correlation**

By:

Mohammad Faraji

Supervisor:

Dr. Fardin Akhlaghian Tab

Advisor:

Seyed Amjad Seyed

Septamber 2023



**University of Kurdistan
Faculty of Engineering
Department of Computer Software Engineering**

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree
of M.Sc. in Computer Engineering – Artificial Intelligence and Robotics

Title:

**Multi-Label Feature Selection by Exploiting Global and Local
Label Correlation**

By:

Mohammad Faraji

Supervisor(s):

Dr. Fardin Akhlaghian Tab

Advisor:

Seyed Amjad Seyed

September, 2023