



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

Title

**A Graph-Based Semi-Supervised Learning Approach
for Multi-Label Classification**

By

Seyed Amjad Seyedi

Supervisor

Dr. Parham Moradi

Advisor

Dr. Fardin Akhlaghian Tab

**A Thesis Submitted to the Postgraduate Studies
Office in Partial Fulfillment Requirements for the Degree of M.Sc.
in Artificial Intelligence**

Februray 2018



University of Kurdistan
Faculty of Engineering
Department of Computer Software Engineering

Title:
A Graph-Based Semi-Supervised Learning Approach
for Multi-Label Classification

By:
Seyed Amjad Seyedi

Supervisor:
Dr. Parham Moradi

Advisor:
Dr. Fardin Akhlaghian Tab

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

Februray 2018



University of Kurdistan
Faculty of Engineering
Department of Computer Software Engineering

A Thesis Submitted to the Postgraduate Studies Office in Partial
Requirements for the Degree of M.Sc.
in Artificial Intelligence

Title:
**A Graph-Based Semi-Supervised Learning Approach
for Multi-Label Classification**

By:
Seyed Amjad Seyed

The above thesis was evaluated and approved by the following
members of the thesis committee with
excellent quality on Februray 17, 2018.

1. Supervisor: Assoc. Prof. Parham Moradi
2. Advisor: Assist. Prof. Fardin Akhlaghian Tab
3. External Examiner: Assist. Prof. Shahrokh Esmacili
4. Internal Examiner: Assist. Prof. Alireza Abdollahpouri

P. Moradi
F. Akhlaghian TAB
Sh
Abdollahpouri

Head of Department:
Dr. Alireza Abdollahpouri

Abdollahpouri

Faculty Graduate Coordinator:
Dr. Mehrdad Khamforoush

M. Khamforoush

Abstract

Multi-label problems arise in various domains including automatic multimedia data categorization, and have generated significant interest in computer vision and machine learning community. However, existing methods do not adequately address two key challenges: exploiting correlations between labels and making up for the lack of labeled data or even missing labels while exploiting label correlations is important for multi-label learning. Also, existing approaches typically exploit label correlations globally, by assuming that the label correlations are shared by all the instances. In this research a semi-supervised matrix factorization model based on local correlation is proposed to handle these challenges that formulates multi-label learning as a constrained semi-supervised non-negative matrix factorization problem that which allows label correlations to be exploited locally and explicitly introduces an auxiliary basis to learn the semantic subspace and boosts its discriminating ability by exploiting labeled and unlabeled examples together. Particularly, it propagates labels of the labeled examples to the unlabeled one. This method can efficiently deal with large-scale datasets and exploit the label correlations and obtain promising and better label prediction results than state-of-the-art methods.

Keyword: Semi-Supervised Learning, Multi-Label Problems, Graph Regularization, Semi-Nonnegative Matrix Factorization, Local Correlation.

چکیده

مسائل چندبرچسبه در حوزه‌های مختلف از جمله طبقه‌بندی خودکار داده‌های چندرسانه‌ای به وجود آمده‌اند و در زمینه بینایی ماشین و یادگیری ماشین مورد توجه قرار گرفته‌اند. اما روش‌های موجود به دو چالش اصلی، استخراج وابستگی بین برچسب‌ها و کمبود داده‌های برچسب‌گذاری شده و حتی برچسب‌های از دست رفته نمی‌پردازند، درحالی که استخراج همبستگی^۱ برچسب‌ها برای یادگیری چندبرچسبه بسیار مهم است. همچنین رویکردهای فعلی معمولاً با فرض اینکه همبستگی برچسب‌ها در همه موارد به اشتراک گذاشته شده است از همبستگی‌های برچسب سراسری استفاده می‌کنند. لذا در این پژوهش، مدل نیمه نظارتی^۲ تجزیه مبتنی بر همبستگی محلی، برای پوشش چالش‌های ذکر شده پیشنهاد شده است که اجازه می‌دهد همبستگی برچسب‌ها به صورت محلی مورد استفاده قرار داد. این مدل به صراحت پایه کمکی را برای یادگیری زیر فضای معنایی معرفی می‌کند و توانایی‌های تفکیک‌کنندگی آن را با استفاده از نمونه‌های برچسب‌دار و بدون برچسب تقویت می‌کند. به عبارت دیگر، برچسب نمونه‌های برچسب‌دار را به نمونه‌های بدون برچسب منتقل می‌کند. این روش پیشنهادی می‌تواند به طور مؤثر با مجموعه‌های مقیاس بزرگ مقابله و همبستگی برچسب‌ها را استخراج کند و پیش‌بینی‌های امیدوار کننده و بهتری نسبت به روش‌های نوین بدست آورد.

کلمات کلیدی: یادگیری نیمه نظارتی، مسائل چندبرچسبه، منظم‌سازی گراف، تجزیه نیمه -

نامنفی ماتریس، همبستگی محلی

¹ Correlation

² Semi-Supervised

فهرست مطالب

| عنوان | صفحه |
|---|------|
| فصل اول: مقدمه | ۱ |
| ۱-۱- بیان و اهمیت مسئله..... | ۱ |
| ۲-۱- مفروضات پژوهش..... | ۳ |
| ۳-۱- روش تحقیق..... | ۳ |
| ۴-۱- ساختار پایان نامه..... | ۴ |
| فصل دوم: پیشینه‌ی تحقیق | ۵ |
| ۱-۲- مقدمه..... | ۵ |
| ۲-۲- یادگیری نیمه نظارتی..... | ۶ |
| ۱-۲-۱- یادگیری استقرائی و یادگیری انتقالی..... | ۶ |
| ۲-۲-۲- تعریف مسئله نیمه نظارتی..... | ۸ |
| ۳-۲- تجزیه ماتریس..... | ۹ |
| ۱-۳-۲- تجزیه‌های نامنفی مقید ماتریس..... | ۱۱ |
| ۲-۳-۲- تجزیه‌های نامنفی ساختاری ماتریس..... | ۱۳ |
| ۳-۳-۲- تجزیه‌های نامنفی تعمیم یافته ماتریس..... | ۱۴ |
| ۴-۲- یادگیری چندبرچسب..... | ۱۶ |
| ۱-۴-۲- تعریف..... | ۱۶ |
| ۲-۴-۲- چالش..... | ۱۶ |
| ۳-۴-۲- روش‌های یادگیری چندبرچسب..... | ۱۷ |
| فصل سوم: روش پیشنهادی | ۱۹ |
| ۱-۳- مقدمه..... | ۱۹ |
| ۲-۳- روش پیشنهادی اول: تجزیه نیمه نظارتی مبتنی بر همبستگی پویا..... | ۲۰ |
| ۱-۲-۳- ایجاد گراف برچسب..... | ۲۰ |
| ۲-۲-۳- جایگذاری گراف در مسئله (منظم سازی گراف)..... | ۲۱ |
| ۳-۲-۳- گراف همبستگی پویا..... | ۲۲ |
| ۴-۲-۳- تابع هدف پیشنهادی..... | ۲۳ |
| ۵-۲-۳- فرمول‌های به روزرسانی..... | ۲۳ |

| | |
|----|--|
| ۲۵ | ۳-۲-۶- مقداردهی اولیه مبتنی بر اطلاعات نظارتی |
| ۲۷ | ۳-۳- روش پیشنهادی دوم: تجزیه نیمه نظارتی مبتنی بر همبستگی محلی |
| ۲۷ | ۳-۳-۱- همبستگی محلی در تجزیه نیمه-نامنفی ماتریس |
| ۲۹ | ۳-۳-۲- ایجاد گراف‌های برچسب احتمالاتی و خمینه |
| ۳۰ | ۳-۳-۳- منظم‌سازی گراف‌ها |
| ۳۱ | ۳-۳-۴- تابع هدف پیشنهادی |
| ۳۲ | ۳-۳-۵- فرمول‌های بروز رسانی |
| ۳۳ | ۳-۳-۶- مقداردهی اولیه مبتنی بر اطلاعات نظارتی و خوشه‌بندی |
| ۳۵ | فصل چهارم: نتایج و تفسیر آن‌ها |
| ۳۵ | ۴-۱- مقدمه |
| ۳۶ | ۴-۲- مجموعه داده‌ها |
| ۳۶ | ۴-۲-۱- مجموعه داده‌های چندکلاسه |
| ۳۷ | ۴-۲-۲- مجموعه داده‌های چندبرچسبه |
| ۳۸ | ۴-۳- معیارهای ارزیابی |
| ۳۸ | ۴-۳-۱- معیارهای ارزیابی مسائل چندکلاسه |
| ۳۹ | ۴-۳-۲- معیارهای ارزیابی مسائل چندبرچسبه |
| ۴۰ | ۴-۴- نتایج |
| ۴۰ | ۴-۴-۱- نتایج روش پیشنهادی اول |
| ۴۲ | ۴-۴-۲- نتایج روش پیشنهادی دوم |
| ۴۶ | فصل پنجم: نتیجه‌گیری و پیشنهادات |
| ۴۶ | ۵-۱- نتیجه‌گیری |
| ۴۷ | ۵-۲- پیشنهادات برای تحقیقات آتی |
| ۴۸ | فهرست منابع |
| ۵۳ | پیوست |

فهرست اشکال

| عنوان | صفحه |
|--|------|
| شکل ۱-۲- مثالی از یادگیری استقرائی روی نمونه‌های آموزشی | ۷ |
| شکل ۲-۲- مثالی از یادگیری استقرائی بعد از اضافه کردن نمونه‌های آزمایشی | ۷ |
| شکل ۳-۲- مثالی از یادگیری نیمه‌نظارتی مبتنی بر گراف | ۸ |
| شکل ۴-۲- انواع تجزیه‌های نامنفی ماتریس | ۱۱ |
| شکل ۱-۳- نمودار روش پیشنهادی اول | ۲۳ |
| شکل ۲-۳- ساختار سلسله مراتبی تجزیه | ۲۸ |
| شکل ۳-۳- الف) مفهوم خمینه در فضای داده ب) مفهوم نمونه در فضای برچسب | ۳۱ |
| شکل ۴-۳- نمودار مقداردهی اولیه ماتریس‌های | ۳۴ |
| شکل ۱-۴- ارزیابی روش پیشنهادی اول بر روی مجموعه داده‌ها براساس معیار ارزیابی F-measure | ۴۱ |
| شکل ۲-۴- ارزیابی روش پیشنهادی اول بر روی مجموعه داده‌ها براساس مدت زمان اجرا | ۴۲ |
| شکل ۳-۴- ارزیابی روش پیشنهادی دوم بر روی مجموعه داده Corel5K براساس معیار میانگین درستی | ۴۳ |
| شکل ۴-۴- ارزیابی روش پیشنهادی دوم بر روی مجموعه داده Corel5K براساس معیار خطای رتبه‌بندی | ۴۴ |
| شکل ۵-۴- ارزیابی روش پیشنهادی دوم بر روی مجموعه داده Enron براساس معیار میانگین درستی | ۴۵ |
| شکل ۶-۴- ارزیابی روش پیشنهادی دوم بر روی مجموعه داده Enron براساس معیار خطای رتبه‌بندی | ۴۵ |

فهرست جداول

| عنوان | صفحه |
|--|------|
| جدول ۴-۱- مجموعه داده‌های استفاده شده در روش پیشنهادی اول..... | ۳۷ |
| جدول ۴-۲- مجموعه داده‌های استفاده شده در روش پیشنهادی دوم..... | ۳۷ |

جدول نمادها*

| نماد | شرح لاتین | شرح |
|---------------|---|-----------------------------------|
| X | data matrix | ماتریس داده |
| d | number of features | تعداد ویژگی‌ها |
| n | number of samples | تعداد نمونه‌ها |
| Y | label matrix | ماتریس برچسب |
| l | number of label | تعداد برچسب |
| Z | label centers | ماتریس مراکز برچسب |
| H | label membership | ماتریس تعلقات برچسب |
| W | semantic relevance graph | گراف ارتباط معنایی |
| μ | average of Gaussian kernel | میانگین هسته گاوسی |
| σ | variance of Gaussian kernel | واریانس هسته گاوسی |
| D^W | degree matrix of semantic relevance | ماتریس درجه گراف ارتباط معنایی |
| L | Laplacian matrix of semantic relevance | ماتریس لاپلاسی گراف ارتباط معنایی |
| D | pairwise distance matrix | ماتریس فاصله دویه‌دو |
| λ | semantic relevance regularization parameter | پارامتر تنظیم ارتباط معنایی |
| P | Transition matrix | ماتریس گذار |
| nl | number of labeled samples | تعداد نمونه‌های برچسب‌دار |
| nu | number of unlabeled samples | تعداد نمونه‌های بدون چسب‌دار |
| m | number of categories | تعداد دسته‌ها |
| A | category centers | ماتریس مراکز دسته |
| Q | category membership | ماتریس تعلقات دسته |
| C | correlation graph | گراف همبستگی |
| k | number of neighbors | تعداد همسایه |
| G | manifold graph | گراف خمینه |
| D^G | degree matrix of manifold | ماتریس درجه خمینه |
| S | Laplacian matrix of manifold | ماتریس لاپلاسی خمینه |
| λ_1 | sparsity regularization parameter | پارامتر تنظیم تُنکی |
| λ_2 | manifold regularization parameter | پارامتر تنظیم خمینه |
| λ_3 | semantic relevance regularization parameter | پارامتر تنظیم ارتباط معنایی |
| V | sparsity factor | عامل تُنکی |
| ε | very small constant | ثابت بسیار کوچک |

* حروف بزرگ انگلیسی بیانگر ماتریس و حروف کوچک انگلیسی بیانگر بردار یا عدد هست.

فصل اول: مقدمه

۱-۱- بیان و اهمیت مسئله

با رشد سریع محتوای برخط از قبیل تصویر، ویدئو، صفحات وب و غیره، طراحی یک سیستم طبقه‌بند مقیاس‌پذیر و مؤثر برای مدیریت، ذخیره و جستجوی محتوا، حیاتی است. در طبقه‌بندی مرسوم، فرض می‌شود هر نمونه به صورت کامل به یک کلاس از مجموعه‌ای محدود از کلاس‌ها، تعلق دارد. هرچند در کاربردهای نوین، یک نمونه می‌تواند برچسب‌های مختلفی داشته باشد. برای مثال، در طبقه‌بندی معنایی منظره، یک تصویر می‌تواند به وسیله چندین برچسب ادراکی، برچسب-گذاری شود. بنابراین، با توجه به این که داده‌های چندبرچسب در بسیاری از حوزه‌های کاربردی (مانند بازیابی اطلاعات چند رسانه‌ای، پیشنهاد تگ، طبقه‌بندی پرس و جو، پیش‌بینی عملکرد ژن، تشخیص پزشکی، کشف مواد مخدر و بازیابی) به طور گسترده‌ای وجود دارند، رسیدگی به چنین وظایفی، مبحث یادگیری چندبرچسب را به وجود آورده که در سال‌های اخیر توجه زیادی را به خود جلب نموده است.

یک راه حل ساده برای یادگیری چندبرچسب این است که مسئله به مجموعه‌ای از طبقه‌بندهای دودویی تجزیه شود. با این حال، در این روش این واقعیت نادیده گرفته می‌شود که اطلاعات یک برچسب ممکن است برای یادگیری یک برچسب مرتبط دیگر مفید باشد؛ به ویژه در مواقعی که برای برخی از برچسب‌ها، نمونه‌های آموزشی کافی وجود نداشته باشد، همبستگی برچسب ممکن است اطلاعات مفیدی را ارائه دهد. بنابراین، همبستگی برچسب به طور کلی به عنوان یک مؤلفه کلیدی از رویکرد یادگیری چند برچسب پذیرفته شده است [۱-۳].

نظر به اینکه دانش قبلی اغلب در کاربردهای واقعی در دسترس نیست، بسیاری از رویکردها [۴-۶] سعی می‌کنند تا براساس داده‌های آموزش، برچسب‌ها را یاد بگیرند و همبستگی برچسب‌ها را در فرآیند یادگیری درک کنند. علاوه بر این، برای یادگیری چندبرچسبه چندین کار با استفاده از تکنیک‌های دیگر یادگیری ماشین، مانند یادگیری چندبرچسبه-چندنمونه‌ای [۷]، یادگیری فعال^۱ برای یادگیری چندبرچسبه [۸] و یادگیری چندبرچسبه همراه با یادگیری چندهسته‌ای [۹] انجام شده‌اند.

الگوریتم‌های فوق‌الذکر با تمام برچسب‌ها به طور یکسان رفتار می‌کنند و همچنین در هنگام آموزش مدل‌ها، تمام نمونه‌های آموزش را در هر برچسب به صورت یکسان در نظر می‌گیرند. با این حال، در دنیای واقعی، پیچیدگی‌های مختلف آموزش ممکن است وجود داشته باشد و پیچیدگی‌های مختلف آموزش در یک برچسب نیز متفاوت است. برای مثال، ممکن است دو تصویر که دارای یک برچسب مشابه باشند میزان تعلق به آن کلاس‌ها متفاوت باشد

هدف طبقه‌بندی چندبرچسبه، یادگیری فضای ویژگی با استفاده از فضای برچسب‌ها است، بنابراین، هر دو اطلاعات ویژگی و اطلاعات برچسب باید برای بهره‌برداری از همبستگی برچسب استفاده شود. با این حال، این روش‌های موجود ممکن است زمان‌بر باشند [۱۰، ۱۱] که عمدتاً به اندازه فضای پنهان از پیش تعریف شده بستگی دارد [۱۲-۱۴].

از سوی دیگر، روش‌های یادگیری نظارتی، طبقه‌بندها را بر اساس نمونه‌های به اندازه و چند برچسب دار آموزش می‌دهند. با این حال، این روش برای به دست آوردن موارد چندمنظوره بسیار هزینه‌بر است. در عمل، ما همیشه با تعداد کمی از نمونه‌های دارای چند برچسب و مقدار زیادی نمونه‌های بدون برچسب مواجه هستیم. بنابراین، بسیار مهم است که روش‌های یادگیری چندبرچسبه نیمه‌نظارتی داشته باشیم که می‌توانند از هر دو نوع داده چندبرچسبه و داده‌های بدون برچسب برای حل این مشکل مهم استفاده کنند. روش‌هایی وجود دارد که در آنها تکنیک‌های یادگیری فعال یکپارچه‌ای در فرایند یادگیری چندبرچسبه برای بهبود عملکرد پیش‌بینی شده است. درواقع، یک ماتریس مشترک ویژگی همراه با یکپارچه‌سازی داده‌های چندبرچسبه و بدون برچسب ساخته شده است. گام اصلی الگوریتم تکرار آن‌ها این است که تجزیه مقدار منحصر به فرد ماتریس مشترک را محاسبه کنیم که این گونه روش‌ها در مواقعی که تعداد اشیاء، ویژگی‌ها و برچسب‌ها زیاد هستند هزینه محاسباتی بالایی را به وجود می‌آورد.

¹ Active Learning

سهم اصلی این پژوهش، توسعه یک روش جدید برای یادگیری چند برچسب است که تنها تعداد کمی از داده های چند برچسب دار وجود دارد. ایده اصلی ما که در آن مفهوم تجزیه ماتریس نیز استفاده شده است، این است که تجزیه نامنفی ماتریس از الگوریتم های تجزیه و تحلیل چندمتغیره و جبر خطی است که در آن ماتریس X به دو ماتریس Z و H تجزیه شده است، با این ویژگی که یک ماتریس عنصر منفی ندارد. این نامنفی بودن و همچنین شکسته شدن به رتبه های کوچکتر، ماتریس های حاصل را برای محاسبه آسان تر می کند، که به طور گسترده ای برای یادگیری ماشین (عمدتاً روش خوشه بندی) کاربردی تر هستند و با برخی تغییرات در چارچوب تجزیه ماتریس قابل انطباق می باشند.

۱-۲- مفروضات پژوهش

- فرض می شود هر نوع داده (تصویر، صوت، متن و ...) به صورت یک بردار ویژگی بازنمایی شده است و مجموعه داده به صورت داده های ساختاری^۱ می باشد.
- تعداد محدودی از داده ها به صورت برچسب دار به مسئله داده می شود و بیشتر داده ها بدون برچسب هستند.
- به دلیل انتقالی بودن یادگیری، کلیه داده ها به صورت یکجا وارد الگوریتم می شود و فازهای آموزش و آزمایش در این روش وجود ندارد.
- برچسب داده چندبرچسبه به صورت برداری دودویی بیان شده و هدف این یادگیری، اختصاص دادن چند برچسب محتمل به نمونه های بدون برچسب می باشد.

۱-۳- روش تحقیق

در ابتدا، منابع و مقالات موجود در حوزه یادگیری نیمه نظارتی مورد بررسی قرار گرفت. با توجه به مطالعات و بررسی عملکرد روش های مختلف به این نتیجه رسیدیم که در مقالات اخیر از روش های مبتنی بر تجزیه ماتریس برای حل مسائل طبقه بندی، استفاده شده است. از میان روش های مبتنی بر تجزیه، تجزیه نیمه نامنفی ماتریس قدرت تفکیک و همگرایی مناسب تری دارد. سپس بر روی مسائل چند برچسبه مطالعه صورت گرفت و انواع تعریف همبستگی در این مسائل ارزیابی شد. با توجه به این که اغلب روش های مبتنی بر همبستگی مرتبه پایین و همبستگی سراسری بودن

¹ Structured data

سعی شد از همبستگی‌های مرتبه بالا و نوین در پژوهش کنونی استفاده شود با مطالعه در روش‌های مرتبه بالا مناسب‌ترین همبستگی انتخاب شد. درنهایت مجموعه داده‌ها، کتابخانه‌ها و معیارهای ارزیابی شناخته شده مسائل چندبرچسب شناسایی شده و نحوه کار با آنها بررسی شد.

۱-۴- ساختار پایان‌نامه

متن این پایان‌نامه با احتساب فصل مقدمه در پنج فصل نگارش شده است. خلاصه‌ی رئوس مطالب اصلی هر فصل به شرح زیر است.

در فصل ۱ (که فصل جاریست)، مقدمه‌ای بر پایان‌نامه شامل تعریف مسئله، تاریخچه و اهمیت آن، هدف از پژوهش، مفروضات و روش انجام پژوهش آورده شده است.

در فصل ۲، ابتدا مفاهیم پایه‌ی یادگیری نیمه‌نظارتی، تجزیه ماتریس و مسئله‌ی طبقه‌بندی چندبرچسب بیان می‌شود، سپس روش‌های انجام‌شده در زمینه‌ها دسته‌بندی‌شده، و هر کدام از دسته‌ها به تفصیل بررسی می‌شود، و مزایا و معایب هر کدام بیان می‌شود.

در فصل ۳، ابتدا روش پیشنهادی اول برای مسائل چندکلاسه با جزئیات معرفی می‌شود. بازنمایی مسئله، جایگذاری اطلاعات نیمه‌نظارتی، تابع هدف استفاده‌شده و معایب این روش، به ترتیب توضیح داده می‌شوند. سپس روش پیشنهادی دوم در جهت بهبود نتایج روش اول با ذکر جزئیات شرح داده می‌شود.

در فصل ۴، شرایط آزمایش، مجموعه داده‌های مورد استفاده، و معیارهای ارزیابی بیان می‌شوند. سپس، نتایج آزمایش‌ها روش‌های پیشنهادی اول و دوم بر روی مجموعه داده‌ها، در دو بخش مجزا ارائه می‌شود.

در فصل ۵، به جمع‌بندی کلی از روش‌های ارائه‌شده در این پایان‌نامه پرداخته می‌شود و پیشنهادهایی برای توسعه‌ی روش‌های ارائه‌شده برای کارهای آینده، ارائه می‌شود.

فصل دوم: پیشینه‌ی تحقیق

۲-۱- مقدمه

با رشد روزافزون داده‌ها و اهمیت تحلیل داده، شاهد استفاده از الگوریتم‌های یادگیری ماشین در حوزه‌های مختلف هستیم. وب با حجم عظیمی از اطلاعات خام، بستری مناسب برای تحلیل داده می‌باشد، و مملو از داده‌هایی چندماهیتی است که اطلاعی از محتوای آن‌ها وجود ندارد. ارائه راه حلی که بتواند با توجه به محدود بودن نمونه‌های برچسب‌دار، داده‌های خام را از منظرهای گوناگون تفسیر کرده و ماهیت آن‌ها را تشخیص دهد می‌تواند مفید باشد. برای ارائه راه حلی با توانایی ذکر شده لازم است در این زمینه از روش‌ها و الگوریتم‌های یادگیری به‌روز شناخت مناسبی داشت. در این فصل، به منظور آشنایی با مفاهیم مطرح شده در این پایان نامه، سعی می‌شود مفاهیم اولیه از جمله: یادگیری نیمه‌نظارتی، یادگیری‌های استقرائی و انتقالی، مسئله تجزیه ماتریس و یادگیری چندبرچسبه، شرح داده شود. سپس روش‌های مطالعه شده در این پژوهش، از قبیل روش‌های شناخته‌شده و روش‌های نوین در زمینه‌های تجزیه ماتریس و یادگیری چندبرچسبه معرفی خواهد شد.

۲-۲- یادگیری نیمه نظارتی

در تشخیص الگوی سنتی، یادگیری در نمونه‌های بدون ناظر (که در آن همه داده‌ها بدون برچسب هستند) یا در نمونه‌های باناظر (که همه داده‌ها برچسب دارند)، صورت گرفته است. اما یادگیری نیمه نظارتی روشی است که هم از داده‌های برچسب‌دار و هم از داده‌های بدون برچسب استفاده می‌کند [۱۵]. هدف از یادگیری نیمه نظارتی این است که با ترکیب داده‌های برچسب‌دار و غیربرچسب‌دار رفتار یادگیری را بهبود بخشید، و الگوریتم‌هایی را طراحی کرد که از فواید این ترکیب‌ها استفاده کنند [۱۶]. این نوع یادگیری در یادگیری ماشین کاربردهای فراوانی مانند: طبقه‌بندی متن^۱، حاشیه نویسی تصویر^۲ و برچسب‌زنی اجزای کلام^۳ دارد زیرا به آسانی می‌توان با داده‌های بدون برچسب در دسترس، کارایی یادگیری باناظر را با توجه به عدم دسترسی و گران بودن داده‌های برچسب‌دار، بالا برد.

"یادگیری نیمه نظارتی از نمونه‌های برچسب‌دار و بدون برچسب برای انجام یک یادگیری نظارتی و یا بدون نظارت، استفاده می‌کند و می‌توان گفت این روش، روشی مابین یادگیری نظارتی و یادگیری بدون نظارت می‌باشد [۱۵]."

۲-۲-۱- یادگیری استقرائی و یادگیری انتقالی

برای درک بهتر یادگیری نیمه نظارتی در این بخش سعی می‌شود، انواع یادگیری از لحاظ نحوه استفاده از داده‌های مسئله (داده‌های برچسب‌دار و بدون برچسب) را به صورت اجمالی توضیح داد.

- یادگیری استقرائی^۴: رویکرد استقرائی برای حل مسائل بدین صورت است که در فاز آموزش از نمونه‌های موجود^۵ (برچسب‌دار) برای آموزش یک الگوریتم یادگیری نظارتی استفاده می‌شود و سپس در فاز آزمایش، برچسب نمونه‌های دیده نشده^۶ (بدون برچسب) پیش‌بینی می‌شود. به زبان آماری می‌توان گفت: "استفاده از داده گذشته برای پیش‌بینی آینده".

^۱ Text classification

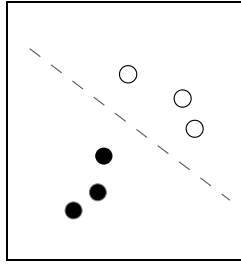
^۲ Image annotation

^۳ Part-of-speech tagging

^۴ Inductive learning

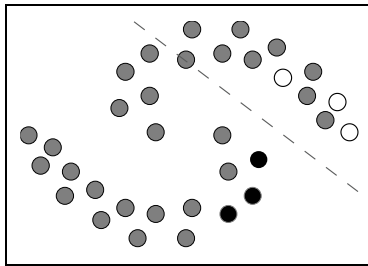
^۵ In-sample

^۶ Out-of-sample



شکل ۱-۲- مثالی از یادگیری استقرائی روی نمونه‌های آموزشی

- یادگیری انتقالی^۱: رویکرد انتقالی یا ترارسانی، در تضاد با رویکرد استقرا، سعی دارد از نمونه‌های برچسب‌دار و بدون برچسب به نحوی استفاده کند که بتواند برچسب نمونه‌های بدون برچسب را پیش‌بینی نماید. هرچند این روش در پیش‌بینی برچسب نمونه‌های بدون برچسب به خوبی عمل می‌کند اما توانایی پیش‌بینی نمونه‌های دیده نشده جدید را ندارد [۱۷].



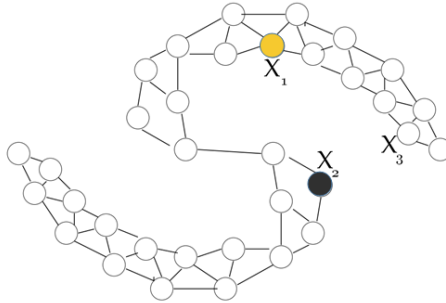
شکل ۲-۲- مثالی از یادگیری استقرائی بعد از اضافه کردن نمونه‌های آزمایشی

برای بیان ساده‌تر مشکل یادگیری استقرائی یک مثال ساده مطرح شده است. همانطور که در شکل ۱-۲ مشخص است نمونه‌های برچسب‌دار (آموزشی) پس از طبقه‌بندی به سادگی قابل تفکیک می‌باشند اما در مسائل نظارتی و به‌خصوص نیمه‌نظارتی، همیشه نمونه‌های آموزشی به صورت یکنواخت در کل فضای کلاس توزیع نشده‌اند. با توجه به شکل ۲-۲ همانطور که مشاهده می‌شود با اضافه کردن داده‌های بدون برچسب (آزمایشی) به فضای مسئله، مدل حاصل دارای خطا طبقه‌بندی می‌باشد. بنابراین اگر در هنگام یادگیری، از ساختار داده‌های بدون برچسب هم استفاده شود می‌توان برچسب این نمونه‌ها را با دقت بهتری پیش‌بینی کرد.

^۱ Transductive learning

۲-۲-۲- تعریف مسئله نیمه نظارتی

در این مسئله، فرض می‌شود فضای ورودی یک ماتریس $X \in \mathbb{R}^{d \times n}$ است به صورت $X = \{x_1, x_2, \dots, x_n\}$ که شامل n داده برداری است که هر کدام d ویژگی دارند. همچنین فضای برچسب یک ماتریس $Y \in \{0, 1\}^{l \times nl}$ است که هر بردار y_i متناظر با یک بردار ورودی x_i می‌باشد. داده‌های برچسب به صورت $\{(x_i, y_i)\}_{i=1}^{nl}$ و داده‌های بدون برچسب را به صورت $\{x_i\}_{i=nl+1}^{nl+nu}$ نشان داده می‌شود که nl تعداد برچسب‌دار و nu تعداد عناصر بدون برچسب مسئله است. هدف پیدا کردن یک نگاشت $f: X \rightarrow Y$ است که $f \in \mathcal{F}$ و \mathcal{F} فضای فرضیه می‌باشد.



شکل ۲-۳- مثالی از یادگیری نیمه نظارتی مبتنی بر گراف

- هر روش نیمه نظارتی باید یک یا چند فرض از فرض‌های زیر را پوشش دهد:
- فرض همواری^۱: اگر دو نقطه در یک منطقه با چگالی بالا نزدیک به هم باشند، برچسب‌های متناظر آن‌ها هم باید نزدیک باشند.
- فرض خوشه^۲: داده‌های موجود در یک خوشه بایستی برچسب‌های مشابهی داشته باشند.
- فرض خمینه^۳: در فضا با بُعد بالا داده‌ها تقریباً روی یک خمینه با بُعد پایین تر قرار دارند و تابع جداساز روی خمینه هموار است [۱۵].

¹ Smoothness assumption

² Clustering assumption

³ Manifold assumption

۲-۳- تجزیه ماتریس

تجزیه ماتریس یکی از موضوعات جبرخطی عددی است. در طول چندین دهه انواع مختلفی از الگوریتم‌های تجزیه ماتریس برای عملیات ماتریس از قبیل حل سیستم‌های خطی، تجزیه طیفی و شناسایی زیرفضا، توسعه یافته‌اند. متمر ثمر بودن برخی از این الگوریتم‌ها در آمار و تحلیل داده اثبات شده است قابل توجه‌ترین آنها تجزیه مقدار منفرد^۱ است که اساس کار تجزیه و تحلیل مؤلفه اصلی^۲ می‌باشد [۱۸].

کارهای اخیر در حوزه یادگیری ماشین بر روی تجزیه‌های ماتریس تمرکز کرده‌اند که به طور مستقیم برخی از ویژگی‌های تحلیل آماری را هدف قرار داده‌اند. به طور خاص، تجزیه نامنفی ماتریس بر روی تحلیل ماتریس‌های داده که عناصر نامنفی دارند (مانند مجموعه داده‌های متن و تصویر)، تمرکز می‌کند. علاوه بر این، تجزیه نامنفی منجر به ایجاد ماتریس‌های نامنفی می‌شود که می‌تواند از دیدگاه تفسیر داده، کارآمد باشد [۱۹].

در سال‌های اخیر دامنه تحقیق تجزیه نامنفی به سرعت گسترده شده است. تجزیه نامنفی در انواع زمینه‌های کاربردی از قبیل محیط زیست، شیمی آماری، شناسایی الگو، تحلیل داده چندرسانه‌ای، متن کاوی، تحلیل بیان ژن^۳ و تعامل پروتئین^۴ قابل استفاده بوده است. توسعه‌های الگوریتمی تجزیه نامنفی برای تطبیق با انواع توابع هدف و انواع مسائل تحلیل داده از قبیل طبقه‌بندی و فیلترسازی مشارکتی^۵ ارائه شده‌اند. برخی مطالعات بر روی توسعه روش‌های محاسباتی برای تجزیه نامنفی صورت پذیرفت و سرانجام پژوهشگران با استفاده از تابع هدف کمترین مربعات، شروع به کاوش در ارتباطات مابین تجزیه نامنفی و خوشه‌بندی k-means کردند [۲۰].

یک بردار نامنفی تصادفی x با d بُعد را فرض کنید که n مشاهده را می‌توان به صورت ماتریس نامنفی $X = [x_1, x_2, \dots, x_n] \in \mathbb{R}_{d \times n}^+$ نشان داد. هدف تجزیه نامنفی تجزیه X به ماتریس پایه نامنفی $Z = [z_1, z_2, \dots, z_l] \in \mathbb{R}_{d \times l}^+$ و ماتریس ضرایب نامنفی $H = [h_1, h_2, \dots, h_n] \in \mathbb{R}_{l \times n}^+$ است به نحوی که:

$$X \approx \sum_{i=1}^l z_i h_i. \quad (1-2)$$

¹ Singular value decomposition

² Principal component analysis

³ Gene expression analysis

⁴ Protein interaction

⁵ Collaborative filtering

که در آن v_i وزن ضرایب مشاهده x_i روی ستون‌های Z است و Z بردارهای پایه یا بردارهای خصوصیات پنهان ماتریس X می‌باشد. به عبارت دیگر، این رویکرد تلاش می‌کند به وسیله ابعاد بسیار کوچک، الگوی تصادفی با ابعاد بزرگ را نمایش دهد. بنابراین اگر ویژگی‌های ذاتی در Z مشخص شوند، تقریبی کامل حاصل می‌شود [۲۱].

در اغلب موارد، تجزیه نامنفی به عنوان یک ابزار کاهش ابعاد یا تکنیک استخراج ویژگی و به صورت $d \ll l$ و $n \ll l$ استفاده می‌شود. هرچند در واقع l می‌تواند بزرگتر یا مساوی d نیز باشد اما در تجزیه $d < l$ و $l > d$ تفاوت‌های اساسی وجود دارد. به عنوان مثال در کدگذاری تُنک^۱ مقدار l از d بزرگتر بوده و لازم نیست مقدار l به ابعاد ماتریس X محدود شده باشد. در این حالت از تُنکی (به دلیل نامنفی بودن و ارائه مازاد) استفاده می‌شود و این برای برخی کاربردها مانند طبقه‌بندی و انتخاب ویژگی، مناسب می‌باشد.

هدف اصلی تجزیه نامنفی ماتریس، تلاش برای پیدا کردن راه حل‌های کارآمد و مؤثر برای این مسئله تحت محدودیت نامنفی است که اساس امکان‌پذیری تجزیه نامنفی را فراهم می‌کند. با توجه به سخت بودن مسئله (سختی زمان چندجمله‌ای غیرقطعی^۲) و فقدان فرمول‌سازی محدب^۳ قطعی، به طور معمول فرمول‌ساز غیرمحدب با قابلیت حل نسبتاً آسان اتخاذ می‌شود و در این شرایط فقط رسیدن به بهینه محلی با هزینه محاسباتی معقول قابل دسترس می‌باشد. از این رو قدیمی‌ترین و عملی‌ترین رویکرد برای این کار، اجرای کمینه‌سازی متناوب یک تابع هدف مناسب می‌باشد؛ به نحوی که، معیار شباهت مابین ماتریس X و حاصل ضرب ZH باشد. مدل‌های مختلف بهینه‌سازی، به طور عمده در توابع هدف و روش‌های بهینه‌سازی متفاوت هستند. مربع فاصله اقلیدسی رایج‌ترین تابع هدف مورد استفاده است [۲۲]:

$$D_F(X \parallel ZH) = \min_{U,V} \frac{1}{2} \|X - ZH\|_F^2 = \frac{1}{2} \sum_{ij} (x_{ij} - [ZH]_{ij})^2 \quad (2-2)$$

که در این فرمول $\|\bullet\|_F$ نشان‌دهنده نُرم فروبنیوس^۴ می‌باشد.

با وجود گذشت دو دهه از به کارگیری تجزیه‌های ماتریس در زمینه یادگیری ماشین، الگوریتم‌های مختلفی ارائه شده‌اند [۲۳، ۲۴]. با توجه به کاربردهای مختلف یادگیری ماشین، محققان سعی

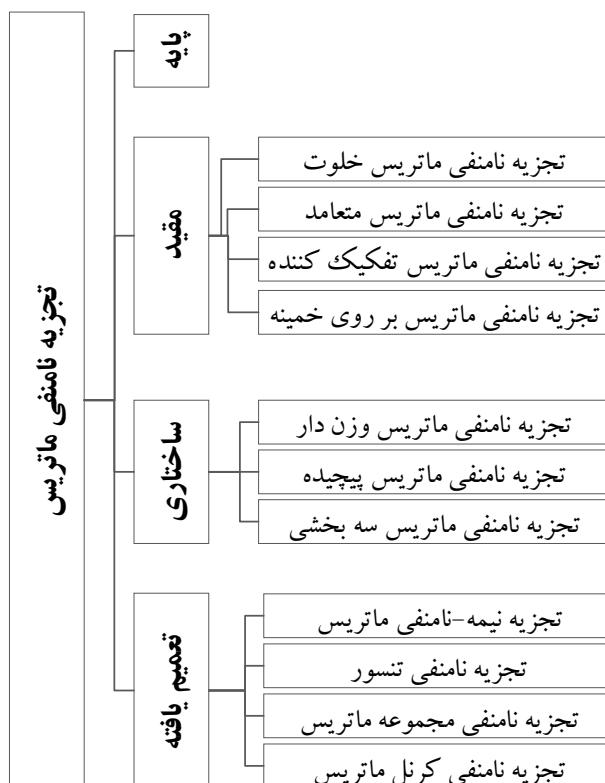
¹ Sparse coding

² NP-hardness

³ Convex

⁴ Frobenius norm

کرده‌اند مسئله تجزیه را با متناسب با کاربرد موردنظر خود تغییر دهند، که هر کدام ویژگی‌ها و قابلیت‌های مختص خود را دارند. در این بخش انواع الگوریتم‌های تجزیه معرفی می‌شوند که شامل: تجزیه نامنفی مقید ماتریس، تجزیه‌های نامنفی ساختاری ماتریس، تجزیه‌های نامنفی ساختاری ماتریس، و تجزیه‌های نامنفی تعمیم‌یافته ماتریس هستند [۲۰].



شکل ۲-۴- انواع تجزیه‌های نامنفی ماتریس

۲-۳-۱- تجزیه‌های نامنفی مقید ماتریس

تجزیه ماتریس پایه فقط با قید نامنفی، راه حل یکتایی ندارد. برای رفع این نقص، لازم است به عنوان جمله منظم‌ساز، یک قید کمکی اضافه بر روی ماتریس‌ها اعمال شود که دانش قبلی را مشارکت می‌دهد و خصوصیات جامع‌تری از موضوع را منعکس می‌کند. انواع مختلفی از مدل‌های تجزیه مقید می‌توانند به صورت یک تابع هدف توسعه یافته واحد تعریف شوند.

$$D_c(X \parallel ZH) = D(X \parallel ZH) + \alpha J_1(Z) + \beta J_2(H) \quad (3-2)$$

که $J_1(Z)$ و $J_2(H)$ جمله‌های جریمه برای وادار کردن قیدهای وابسته به کاربرد موردنظر هستند و α و β پارامترهای منظم‌سازی هستند که مابین برازش مناسب و محدودیت‌های مسئله، تعادل ایجاد می‌کنند [۲۵].

تجزیه نامنفی ماتریس تُنک، گسترده‌ترین و عمیق‌ترین روش از انواع تجزیه‌های مقید است و تقریباً در عمل به یک ضرورت تبدیل شده است. شرط تُنکی در بهبود یکتایی تجزیه همراه با اعمال یک بازنمای محلی، مفید است [۲۶]. سوالی که اینجا مطرح می‌شود این است که قید تُنکی بر روی کدام ماتریس اعمال شود که وابسته به نوع کاربرد می‌باشد. اگر بردارهای پایه (ستون‌های U) تُنک هستند، هر بردار پایه فقط بر روی یک بخش کوچک از هر مشاهده x_i تأثیر می‌گذارد. اگر ستون‌های H تُنک باشند، هر مشاهده x_i به وسیله ترکیب خطی از تعداد محدودی از بردارهای پایه تقریب زده می‌شود [۲۷].

تجزیه نامنفی متعامد نوعی دیگر از تجزیه مقید می‌باشد که شرط متعامد بودن را بر روی یکی از ماتریس‌ها اعمال می‌کند. اصل متعامد بودن ابتدا به وسیله [۲۸] و به منظور کاهش دادن افزونگی مابین بردارهای پایه استفاده شد. اگر بردارهای پایه Z متعامد باشند ($Z^T Z = I$) به معنای بیشترین بخش‌های مجزا می‌باشد و اگر سطرهای H متعامد باشند ($HH^T = I$)، متعامد بودن موجب بهبود دقت خوشه‌بندی خواهد شد. همچنین می‌توان قید متعامد را بر روی هر دو ماتریس اعمال نمود که در بیشتر حالات، عملکرد تخمین ضعیفی دارد [۲۹].

از دیدگاه شناسایی الگو، می‌توان تجزیه را به عنوان یادگیری بدون نظارت در نظر گرفت. با ادغام کردن اطلاعات تفکیک‌کننده با تجزیه، تجزیه پایه به روش نظارتی مبدل می‌شود که تجزیه نامنفی ماتریس تفکیک‌کننده [۳۰] یا تجزیه نامنفی ماتریس فیشر [۳۱] نامیده می‌شود. این تجزیه در کاربردهای طبقه‌بندی مانند تشخیص چهره و تشخیص حالت چهره استفاده شده است. در برخی موارد، داده‌های واقعی اغلب از یک زیرخمینه غیرخطی با ابعاد کم، نمونه گرفته شده‌اند، که در محیطی فضا با ابعاد بزرگ قرار گرفته است و به صورت محلی مسطح است و مشابه یک فضای اقلیدسی است. نشان داده شده است که اگر ساختار هندسی ذاتی شناسایی و حفظ شود، عملکرد یادگیری می‌تواند به طور قابل توجهی افزایش یابد [۳۲].

۲-۳-۲- تجزیه‌های نامنفی ساختاری ماتریس

این تجزیه در حل مسائل یادگیری تجزیه نامنفی، خصوصیات یا ساختارهای دیگری را اعمال می‌کند. این روش برخلاف تجزیه مقید، بجای تعریف برخی قیدهای اضافه به عنوان جمله جریمه، به صورت مستقیم فرمول‌سازی تجزیه منظم را اصلاح تغییر می‌دهد و می‌توان آن را به صورت زیر نوشت:

$$X \approx F(ZH) \quad (۴-۲)$$

به صورت مشخص می‌توان تجزیه نامنفی ساختاری را به سه زیرکلاس: تجزیه نامنفی ماتریس وزن‌دار، تجزیه نامنفی ماتریس پیچیده و تجزیه نامنفی ماتریس سه‌بخشی تقسیم کرد فرمول‌سازی‌های وزن‌دار، نسخه‌های تغییر یافته الگوریتم‌های یادگیری هستند، که می‌توانند برای تأکید بر اهمیت نسبی اجزای مختلف استفاده شوند. با معرفی ماتریس وزن W مدل تجزیه نامنفی وزن‌دار به صورت زیر خواهد بود:

$$W \odot X \approx W \odot (ZH) \quad (۵-۲)$$

به‌طور کلی این تجزیه را می‌توان به عنوان حالتی از تقریب کم‌رتبه وزن‌دار^۱ نام برد، که با توجه به وزن‌های از پیش تعریف شده، به دنبال نزدیک‌ترین ماتریس کم‌رتبه به ماتریس ورودی است. اگر ماتریس داده اصلی با برخی درایه‌های گم یا مشاهده‌نشده، ناقص شده باشد [۳۳]، هدف تجزیه پیش‌بینی داده‌های گم شده می‌باشد، که از آن به عنوان تکمیل ماتریس رتبه پایین همراه با نویز یاد می‌شود. این روش به طور قابل توجهی در فیلتر کردن مشارکتی مانند طراحی سیستم‌های توصیه‌گر استفاده می‌شود. برخی مسائل نیز را می‌توان به وسیله اختصاص وزن‌های دودویی به ماتریس داده حل کرد به صورتی که عناصر مشاهده شده مقدار یک و عناصر ناشناخته مقدار صفر می‌گیرند و ماتریس وزن W متناظر ایجاد می‌شود [۳۴].

تجزیه نامنفی ماتریس پیچیده به طور عمده در زمینه جداسازی منبع^۲ کاربرد دارد. به منظور شرکت دادن اطلاعات زمان در مسئله یا به بیانی دیگر، وابستگی بالقوه مابین ستون‌های همسایه ماتریس داده X ، لازم است خصوصیات متغیر زمانی را نیز در نظر گرفت. می‌توان تجزیه پایه را به صورت زیر به تجزیه پیچیده توسعه داد [۳۵]:

^۱ Weighted low-rank approximation

^۲ Source separation

$$X \approx \sum_{i=1}^{T-1} Z_t^{\rightarrow t} H \quad (6-2)$$

که به صورت مجموع حاصل ضرب دنباله‌ای از ماتریس‌های پایه پی‌درپی Z_t و ماتریس‌های ضرایب متناظر $H^{\rightarrow t}$ می‌باشد.

تجزیه نامنفی ماتریس سه‌بخشی نوعی دیگر از تجزیه ساختاری می‌باشد که با ضرب سه ماتریس به صورت $USV \approx X$ ، تجزیه پایه را توسعه می‌دهد [۳۶]. از آنجایی که بدون قید بودن سه ماتریس بی‌معنی می‌باشد، می‌توان آن را به مسئله مقید دو ماتریسی ادغام کرد. هرچند اضافه کردن قید سوم موجب درجه آزادی اضافه می‌شود، بدین ترتیب موجب می‌شود تجزیه خصوصیات جدیدی داشته باشد.

۳-۳-۲- تجزیه‌های نامنفی تعمیم‌یافته ماتریس

تجزیه نامنفی تعمیم‌یافته در مفهوم وسیعی ممکن است به عنوان توسعه تجزیه پایه در نظر گرفته شود. برخلاف تجزیه مقید که برخی قیدهای اضافه به عنوان جمله جریمه معرفی می‌کند، تجزیه تعمیم‌یافته، مدل تجزیه خود را تا حدی مشابه تجزیه ساختاری، توسعه داده است. در این روش‌ها یا الگوی تجزیه عوض می‌شود یا نوع داده تغییر می‌کند و یا قید نامنفی بودن اصلی نقض می‌شود. این نوع تجزیه را می‌توان به چهار زیربخش: تجزیه نامنفی تنسور، تجزیه نامنفی مجموعه ماتریس، تجزیه نامنفی ماتریس کرنل و تجزیه نیمه-نامنفی ماتریس تقسیم کرد.

روش‌های تجزیه مرسوم، داده را به وسیله مرتب کردن آن به صورت یک ماتریس، پیش پردازش می‌کنند که ممکن است ساختار اصلی داده را از بین ببرند. در مقابل این روش‌ها، تجزیه تنسور قرار دارد که تعمیم ذاتی تجزیه ماتریس است. در واقع این نوع تعمیم کم اهمیت نیست چون تجزیه تنسور در مقابل تجزیه نامنفی خصوصیات متنوع بیشتری دارد. در واقع، داده‌ها در مسئله تجزیه به صورت برداری می‌باشند اما در برخی کاربردها داده اصلی به صورت برداری نیست و ممکن است بردارسازی منجر به برخی مشکلات نامطلوب شود [۳۷]. به عنوان مثال بردارسازی تصویری که دوبعدی می‌باشد، موجب از بین رفتن اطلاعات محلی و ساختاری می‌شود. یکی از دلایل اصلی استفاده از تجزیه موضوع یگانگی است، و برای مقابله با ناسازگاری، باید برخی محدودیت‌های شدید را تحمیل کرد [۳۸].

تجزیه مجموعه ماتریس [۳۹] با توجه به این واقعیت که، اگر بردارسازی داده اصلی منجر به تقریب نامطلوب، ضعف عمومیت و بار محاسباتی بالا شود، مسائل یادگیری به مسائل کوچک کم ارزشی تبدیل خواهند شد. این تجزیه به صورت مستقیم بر روی مجموعه ماتریسی که برای پردازش انتخاب شده اند، اجرا می شود. هر ماتریس نمونه به k ماتریس تجزیه می شود که $k-1$ ماتریس عمومی، خصوصیات نهان را به عنوان مجموعه ماتریس خصوصیات نمایش می دهد.

اساساً، تجزیه ماتریس و انواع آن که در فوق ذکر شده اند، مدل های خطی هستند که قادر به استخراج ساختارهای غیر خطی و روابط نهان مابین داده ها نیستند. این امر دامنه کاربرد تجزیه را محدود می کند. برای غلبه بر این محدودیت ها، استفاده از بسط ذاتی با استفاده از روش های مبتنی بر کرنل می باشد که با استفاده از توابع غیر خطی، داده های ورودی را به فضای ویژگی ضمنی نگاشت می کند [۴۰]. علاوه بر این، توانایی این روش این است که در پردازش داده با مقادیر منفی، از برخی از توابع کرنل خاص استفاده می شود و وابستگی های مرتبه بالا بین بردارهای پایه پذیرفته می شود [۴۱]. یک نگاشت غیر خطی را فرض کنید که فضای داده ورودی X را به فضای خصوصیت Y نگاشت می کند. تجزیه نامنفی کرنل سعی می کند ماتریس های V و Z را پیدا کند، به نحوی که $VZ \approx Y$ و H مجموعه بردارهای پایه فضای اصلی می باشد.

تجزیه پایه، ماتریس های پایه و ضرایب و عناصر ماتریس داده X را به نامنفی بودن، محدود می کند. زمانی که ماتریس نامقید باشد می تواند ترکیبی از عناصر مثبت و منفی را داشته باشد. تجزیه نیمه نامنفی همان دیدگاه تجزیه پایه را دارد که مقادیر ماتریس Z همان محدودیت نامنفی بودن را دارند اما ماتریس های داده X و ماتریس H هیچ محدودیت علامتی ندارند. این حالت تعمیم بدین معنی است که در کاربرد عملی، داده ورودی همیشه نامنفی نیست بنابراین ممکن است ویژگی های نهان یا مؤلفه های اصلی همان عناصر نامنفی را داشته باشند که اطلاعات را از همان منظر بازتاب می دهند. در واقع این روش همان تفسیر فیزیکی تجزیه نامنفی را دارد، علاوه بر این ترکیبات غیر کاهشی همچنان مؤثر می باشند. در این روش برای حل مسئله تجزیه از یک رویکرد تکرار متناوب استفاده شده است، که قسمت های مثبت و منفی ماتریس بدون محدودیت را تفکیک می کند. ماتریس Z زمانی که ماتریس H ثابت فرض می شود، به وسیله قوانین ضرب ماتریس برورسانی می شود و سپس با ثابت فرض کردن Z ، راه حل بهینه محلی تحلیلی برای H حاصل خواهد شد. همچنین همگرایی این روش اثبات شده است، علاوه بر این تجزیه نیمه نامنفی معادل مدل پیشین تجزیه محدب [۴۲] به کار برده می شود.

۲-۴- یادگیری چندبرچسبه

۲-۴-۱- تعریف

در طبقه بندی تک برچسبه یک فرض اساسی اتخاذ می شود که هر نمونه واحد، تنها متعلق به یک کلاس می باشد. به عبارت دیگر، هر نمونه تنها یک برچسب یکتا دارد. در حالی که در بسیاری از کاربردهای دنیای واقعی چنین فرض ساده کنند های الزاماً برقرار نیست و اشیاء جهان واقعی می توانند پیچیده شوند و در عین حال چندین برچسب متفاوت داشته باشند.

در این یادگیری، فرض می شود فضای ورودی یک ماتریس $X \in \mathbb{R}^{d \times n}$ است به صورت $X = \{x_1, x_2, \dots, x_n\}$ که شامل n داده برداری است که هر کدام d ویژگی دارند. همچنین فضای برچسب یک ماتریس $Y \in \{0, 1\}^{l \times n}$ است که هر بردار y_i متناظر با یک بردار ورودی x_i است به نحوی که، اگر نمونه x_i ، l امین برچسب را داشته باشد مقدار y_{li} برابر با 1 و در غیر این صورت برابر با 0 خواهد بود.

۲-۴-۲- چالش

چالش اصلی یادگیری از داده های چندبرچسبه، اندازه فضای خروجی است به نحوی که با افزایش تعداد برچسب های کلاس، تعداد مجموعه برچسب ها به صورت نمایی رشد پیدا می کند. برای مقابله با این چالش اندازه نمایی فضای خروجی، ضروریست که فرایند یادگیری با استخراج همبستگی (وابستگی) میان برچسب ها تسهیل شود. به عنوان مثال اگر یک تصویر با برچسب /یران حاشیه نویسی شود احتمال این که برچسب کشتی یا ریزگرد را نیز بگیرد، وجود دارد و یا یک سند اگر برچسب سیاسی بگیرد، گرفتن برچسب سرگرمی آن غیرمحتمل می باشد. از این رو تأثیر استخراج همبستگی برچسب به منظور موفقیت روش های یادگیری چندبرچسبه حیاتی می باشد. راهبردهای موجود برای استخراج همبستگی برچسب را می توان به سه دسته و به عنوان مراتب همبستگی تقسیم کرد [۳]:

راهبرد مرتبه اول: وظیفه این نوع یادگیری چندبرچسبه، به سبک برچسب به برچسب و سپس نادیده گرفتن سایر برچسب ها، مانند تجزیه مسئله یادگیری چندبرچسبه به تعدادی مسئله طبقه بندی دودویی (یک طبقه بند به ازای هر برچسب) است [۴۳، ۴۴]. مزیت اصلی راهبرد مرتبه اول در سادگی مفهومی آن می باشد. از سویی دیگر، به دلیل عدم استفاده از همبستگی برچسب، ممکن است مؤثر بودن روش ها قابل تضمین نمی باشد.

راهبرد مرتبه دوم: وظیفه این نوع یادگیری چندبرچسبه، در نظر گرفتن ارتباط دوه‌دو بین برچسب‌ها، مانند رتبه‌بندی مابین برچسب مرتبط و برچسب نامرتبط، یا تعامل بین هر جفت برچسب است. هنگامی که همبستگی برچسب استخراج می‌شود [۴۵-۴۷]، این روش‌ها، به صورت عمومی کارایی بهتری خواهند داشت. هرچند در کاربردهای واقعی، همبستگی برچسب‌ها از فرض مرتبه دوم فراتر می‌رود.

راهبرد مرتبه بالا: وظیفه این نوع یادگیری چندبرچسبه، در نظر گرفتن ارتباطات مرتبه بالای مابین برچسب‌ها، از قبیل تحمیل کردن تأثیرات سایر برچسب‌ها بر روی هر برچسب یا مشخص کردن ارتباطات میان زیرمجموعه‌های تصادفی برچسب‌ها است [۴۸, ۴۹]. این راهبرد ظرفیت‌های مدل-سازی همبستگی قوی‌تری نسبت به راهبردهای مرتبه اول و مرتبه دوم دارد. از سویی دیگر، پیچیدگی محاسباتی بالاتر و مقیاس‌پذیری کمتری دارد.

۲-۴-۳- روش‌های یادگیری چندبرچسبه

همانطور که در بخش قبل توضیح داده شده است، راهکار محبوب در مورد یادگیری چند برچسبه، این است که، با استفاده از تکنیک‌هایی از قبیل: یک در برابر بقیه^۱، یک در برابر یک^۲ و کدگذاری تصحیح خطا^۳، مسئله را به مجموعه‌ای از مسائل طبقه‌بندی دودویی تقسیم کنیم [۵۰]. این روش‌ها محدودیت‌های خاص خود را دارند، از جمله: مشکل در مقیاس‌پذیری در مجموعه داده‌های بزرگ، عدم توانایی استخراج وابستگی و روابط میان برچسب‌ها به دلیل استفاده از طبقه‌بندهای مستقل و همچنین ممکن است نتایج خروجی‌های طبقه‌بندی هنگامی که تعداد کلاس‌ها زیاد است، نامتعادل شود.

"روش‌هایی که از راهکار مرتبه دوم استفاده می‌کنند از دو نوع همبستگی مابین نمونه‌ها و همبستگی مابین برچسب‌ها استفاده می‌کنند. [۵۱]"

در مقاله [۴۴] روشی ارائه شده است که در آن، احتمالات پیشین برچسب‌ها با استفاده از قانون k نزدیکترین همسایه و اصل بیشینه‌گر احتمال پسین^۴، برای پیش‌بینی برچسب نمونه‌های بدون-برچسب مورد استفاده قرار می‌گیرند. روش انتقالی [۵۲] نیز، ابتدا برچسب نمونه‌های برچسب‌دار را

¹ One-versus-the-rest

² One-versus-one

³ Error-correcting coding

⁴ Maximum A posteriori Principle (MAP)

با محدود کردن اختلاف بین همبستگی برچسب‌ها و شباهت نمونه‌ها بروی نمونه‌های بدون برچسب انتشار می‌دهد، سپس براساس ارتباطات همسایگی نمونه‌ها با راهکار انتشار برچسب به پیش‌بینی برچسب نمونه‌های بدون برچسب می‌پردازد. همچنین روش [۱۱] که یادگیری چندبرچسبه با استفاده از همبستگی محلی نامیده می‌شود، در یک چارچوب واحد یادگیری، برازش متمایز کننده سراسری و حساسیت همبستگی محلی را مشارکت می‌دهد.

علی‌رغم وجود پژوهش گسترده و ارائه مقالات متعدد در زمینه یادگیری چندبرچسبه نظارتی، روش‌های اندکی برای مقابله با مسئله چندبرچسبه نیمه‌نظارتی وجود دارد. یکی از روش‌های محبوب در این زمینه، رتبه‌بندی برچسب [۵۳] است که از طریق برچسب نمونه‌های برچسب‌دار، یک تابع رتبه‌بندی می‌آموزد و هر نمونه بدون برچسب را با آستانه‌سازی امتیازات تابع رتبه‌بندی، طبقه‌بندی می‌کند. هرچند در این روش مقیاس‌پذیری به راحتی صورت می‌پذیرد اما در استخراج همبستگی مابین نمونه‌ها و یا برچسب‌ها عملکرد مناسبی ندارد. از جمله روش‌هایی که از همبستگی برچسب استفاده کرده است روش حداکثر آنتروپی [۴] است که هزینه محاسباتی بالایی دارد. در مقاله [۵۴] یک ساختار سلسله‌مراتبی را برای کنترل اطلاعات همبستگی پیشنهاد داده است. در مقاله [۵۵] یک چارچوب انتشار برچسب همبسته برای یادگیری چندبرچسبه طراحی شده است که اطلاعات کلاس‌های مختلف را با هم ترکیب می‌کند. هرچند نحوه استفاده از همبستگی برچسب مابین نمونه‌های بدون برچسب واضح نیست. مقاله [۴۶] از تجزیه نامنفی مقید استفاده می‌کند تا اطلاعات برچسب را به وسیله این فرض که، الگوهای ورودی مشابه بایستی مجموعه برچسب‌های مشابه داشته باشند، انتشار دهد. همچنین روش [۴۷] با تغییر ساختار تجزیه، بار محاسبات روش قبل را کاهش داده ولی همچنان از همبستگی سراسری مابین برچسب‌ها استفاده کرده که راهکار مناسبی نیست.

فصل سوم: روش های پیشنهادی

۳-۱- مقدمه

در این فصل، ابتدا روش پیشنهادی اول برای مسئله‌ی شناسایی الگو در مجموعه داده‌های دو کلاسه و چند کلاسه توضیح داده می‌شود که از همبستگی سراسری پویا برای بهبود طبقه‌بندی استفاده شده است. این توضیحات در چند بخش تعریف مسئله تجزیه، نحوه ایجاد گراف برجسب، نحوه جایگذاری اطلاعات برجسب در تجزیه، محاسبه همبستگی سراسری و مقداردهی اولیه مقادیر ف ارائه خواهد شد. سپس، روش پیشنهادی اول و معایب آن تحلیل خواهد شد. در ادامه، جزئیات روش پیشنهادی دوم که مبتنی بر همبستگی محلی و توسعه‌یافته روش پیشنهادی اول است برای مسائل چند برجسبه، شرح داده می‌شود؛ شامل بخش‌های استخراج همبستگی محلی، منظم‌سازی خمینه، تجزیه عمیق و تابع هدف نهایی است.

۳-۲- روش پیشنهادی اول: تجزیه نیمه نظارتی مبتنی بر همبستگی پویا

در این بخش، روش پیشنهادی اول به نام تجزیه نیمه نظارتی مبتنی بر همبستگی پویا شرح داده می شود. این الگوریتم از سه فاز عمده تشکیل شده است: در فاز اول شامل مقداردهی اولیه ماتریس ها و ایجاد اطلاعات نیمه نظارتی است، در فاز دوم عمل به روزرسانی ماتریس ها صورت می پذیرد و فاز سوم همبستگی پویا به روزرسانی می شود.

در این روش، فرض می شود فضای ورودی یک ماتریس $X \in \mathbb{R}^{d \times n}$ و فضای برجسب یک ماتریس $Y \in \{0, 1\}^{l \times n}$ است که هر بردار y_i متناظر با یک بردار ورودی x_i است به نحوی که، اگر نمونه x_i ، 1 امین برجسب را داشته باشد مقدار y_{li} برابر با 1 و در غیر این صورت برابر با 0 خواهد بود.

هدف تجزیه از ماتریس X است به نحوی که تعلق نمونه ها به 1 برجسب موجود را به دست آورد.

$$J = \|X - ZH\|_F^2 \quad (1-3)$$

با تجزیه X با در نظر گرفتن 1 برجسب (ویژگی نهان)، دو ماتریس ضرایب $Z_{d \times l}$ و $H_{l \times n}$ حاصل می شود.

۳-۲-۱- ایجاد گراف برجسب

به منظور بازنمایی اطلاعات نظارتی مسئله، گراف W با n گره می سازیم که هر گره نشان دهنده یک نمونه از مجموعه داده اولیه است. گره i به گره j متصل می شود، اگر و فقط اگر یک دانش قبلی وجود داشته باشد که این دو نمونه، برجسب های یکسانی دارند. به عبارت دیگر این دو نمونه از نظر معنایی مشابه می باشند [۵۶].

یال ایجاد شده یک وزن w_{ij} دارد که می توان به روش های مختلفی میزان وزن یا شباهت دو نمونه را مقداردهی کرد. ساده ترین روش استفاده از بازنمایی دودویی است:

$$w_{ij} = \begin{cases} 1 & y_i = y_j \\ 0 & \text{otherwise} \end{cases} \quad (2-3)$$

همچنین می توان ضرب نقطه ای انجام داد:

$$w_{ij} = \begin{cases} x_i^T x_j & y_i = y_j \\ 0 & \text{otherwise} \end{cases} \quad (3-3)$$

که نسبت به (۲-۳) محاسبات بیشتری دارد و یا از تابع کرنل گوسی استفاده کرد:

$$w_{ij} = \begin{cases} \exp - \|x_i - x_j\|^2 / \sigma^2 & y_i = y_j \\ 0 & \text{otherwise} \end{cases} \quad (4-3)$$

که در آن $\|x_i - x_j\|^2$ نشانگر مربع نُرم ۲ اختلاف دو نمونه x_i و x_j و یا همان فاصله اقلیدسی بین این دو نمونه است. همچنین σ ابرپارامتر^۱ این تابع است. در این روش علاوه بر محاسبات اضافه، مشکل تنظیم ابرپارامتر نیز وجود دارد. به منظور سادگی الگوریتم، در این بخش از بازنمایی دودویی (۲-۳) استفاده شده است.

۲-۲-۳- جایگذاری گراف در مسئله (منظم سازی گراف)

در این بخش، به منظور استفاده از اطلاعات نیمه نظارتی، جمله بهینه سازی R معرفی می شود که منجر به همواری^۲ ماتریس تعلقات می شود [۵۷]. این جمله بدین گونه تعریف می شود که، اگر دانش قبلی وجود داشته باشد که نشان دهد دو نمونه از نظر معنایی مشابه هستند پس بایستی این دو نمونه به یک (و یا چند) کلاس یکسان تعلق داشته باشند. به عبارت دیگر، اگر بین دو نمونه i و j مقدار وزن مثبت w_{ij} وجود دارد پس به تناسب این وزن - در فضای برجسب - بردارهای عضویت این دو نمونه باید فاصله کمی باهم داشته باشند.

$$\min_H \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \|h_i - h_j\|^2 w_{ij} = \sum_{j=1}^n h_j^T h_j d_{jj}^W - \sum_{i=1}^n \sum_{j=1}^n h_i^T h_j w_{ij} \quad (3)$$

$$= \text{Tr}(\mathbf{H} \mathbf{D}^W \mathbf{H}^T) - \text{Tr}(\mathbf{H} \mathbf{W} \mathbf{H}^T) = \text{Tr}(\mathbf{H} \mathbf{L} \mathbf{H}^T) \quad (5)$$

در این رابطه h_i و h_j به ترتیب بیانگر بردارهای تعلق نمونه i و j از ماتریس تعلق \mathbf{H} می باشند. ماتریس قطری \mathbf{D}^W ماتریس درجه گراف \mathbf{W} است که از جمع سطری و یا ستونی ماتریس \mathbf{W} حاصل می شود و می توان آن را به صورت $d_{ii}^W = \sum_j w_{ij}$ محاسبه کرد. در ادامه، به منظور سهولت در محاسبه مشتق نرم های ماتریس و استفاده این منظم سازی در تجزیه ماتریس، فرمول (۵-۳) به صورت ماتریسی نوشته شده است. در این فرمول $\text{Tr}(\mathbf{B})$ نشان دهنده اثر ماتریس مربعی \mathbf{B} بوده و برابر حاصل جمع درایه های قطر اصلی آن ماتریس است. همچنین ماتریس \mathbf{L} بیانگر گراف لاپلاسی است، که دانش قبلی ارتباط مابین نمونه ها را ذخیره می کند و به صورت $\mathbf{L} = \mathbf{D} - \mathbf{W}$ تعریف می شود.

¹ Hyperparameter

² Smoothness

۳-۲-۳- گراف همبستگی پویا

به منظور تقویت فرایند یادگیری و به دلیل کم بودن اطلاعات نظارتی، در این روش سعی شده است به گونه‌ای از تعلقات غیرقطعی نمونه‌های بدون برچسب به دست آمده استفاده کرد، که موجب یادگیری دقیق‌تر می‌شود. با توجه به این که تجزیه ماتریس یک فرایند تکراری^۱ است و در هر تکرار ماتریس H که تعلق کلیه نمونه‌ها (نمونه‌های برچسب‌دار و بدون برچسب) به برچسب‌ها را نشان می‌دهد، به روزرسانی می‌شود. مشابه روش‌های *فزون‌سازی*^۲، که یک سیستم یادگیری ضعیف را به سیستم یادگیری قوی تبدیل می‌کند [۵۸] تعلقات را به صورت برچسب‌های نرم^۳ و به عنوان بازخورد^۴، به سیستم یادگیری (فرایند تجزیه ماتریس) وارد می‌کنیم. به منظور تبدیل تعلقات به شباهت برچسب-ها، از مفهوم همبستگی فضای برچسب، مشابه آنچه در الگوریتم *انتشار برچسب پویا* ارائه شده است، استفاده می‌شود:

$$W_{t+1} = H_t^T H_t \quad (۶-۳)$$

که H_t نشان‌دهنده ماتریس تعلقات خروجی حاصل از تجزیه زمان t و W_{t+1} بیانگر همبستگی فضای برچسبی است که به صورت ورودی به فرایند تجزیه زمان $t+1$ بازخورد داده می‌شود.

در فرایند تجزیه ماتریس تعلقات کلیه نمونه‌ها به روزرسانی می‌شود و با توجه به این که نمونه‌های برچسب‌دار، تعلقات قطعی دارند، تغییر این تعلقات، موجب تضعیف اطلاعات نیمه‌نظارتی می‌شود. همچنین در صورت محاسبه همبستگی، میزان همبستگی مابین دو نمونه برچسب‌دار کاهش یافته و اطلاعات نیمه‌نظارتی تأثیری بر منظم‌سازی مسئله نخواهند داشت و بدین گونه مسئله به سمت یادگیری بدون نظارتی سوق پیدا می‌کند. به منظور گریز از این مشکل و اصلاح پویای خطای تجزیه [بر روی نمونه‌های برچسب‌دار]، در این روش مشابه برخی روش‌های *انتقالی*، برچسب نمونه‌های برچسب‌دار به حالت اولیه بازگردانده می‌شود:

$$H^{(l)} = H_0^{(l)} \quad (۷-۳)$$

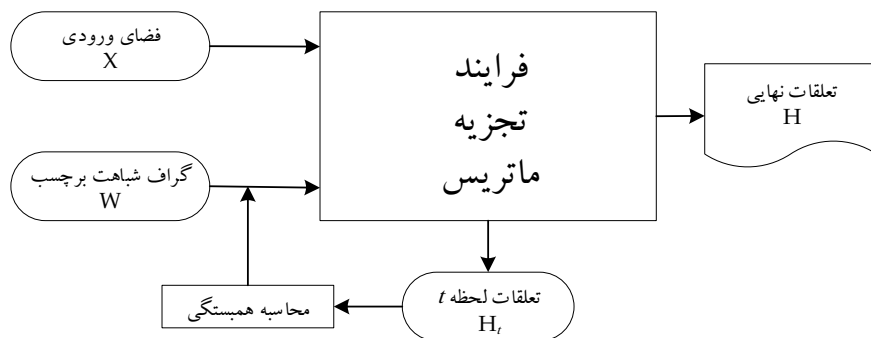
¹ Iterative

² Boosting

³ Soft

⁴ Feedback

که در این فرمول $H^{(l)}$ نشان‌دهنده تعلقات مربوط به نمونه‌های برچسب دار و $H_0^{(l)}$ نشان‌دهنده برچسب نمونه‌های برچسب‌دار (اطلاعات نیمه‌نظارتی اولیه) مسئله می‌باشد. شکل ۳-۱ استخراج همبستگی پویا و نحوه بازخورد آن به فرایند تجزیه ماتریس را نشان می‌دهد.



شکل ۳-۱- نمودار روش پیشنهادی اول

۳-۲-۴- تابع هدف پیشنهادی

همان‌طور که در بخش‌های قبل گفته شد، تابع هزینه تجزیه نیمه-نامنفی ماتریس (۳-۱) رویکردی بدون نظارت بوده و به‌منظور جایگذاری اطلاعات، تابع هزینه نیمه‌نظارتی منظم‌سازی گراف برچسب (۳-۵) ارائه شد. اکنون با اضافه کردن این جمله منظم‌ساز گراف پویا به جمله تجزیه می‌توان ادعا کرد تابع هدف به‌دست آمده یک طبقه‌بندی نیمه‌نظارتی انتقالی می‌باشد. و تابع هدف پیشنهادی با شرط ایجاد گراف همبستگی پویا به‌صورت زیر خواهد بود:

$$\min_{Z, H} \|X - ZH\|_F^2 + \frac{\lambda}{2} \text{Tr}(HLH^T) \quad (۳-۸)$$

$$H \geq 0$$

که λ پارامتر کنترل منظم‌سازی گراف در این بهینه‌سازی می‌باشد و تنها محدودیت این بهینه‌سازی نامنفی بودن مقادیر ماتریس H می‌باشد.

۳-۲-۵- فرمول‌های به‌روزرسانی

هرچند باوجود غیرمحدب بودن اکثر روش‌های تجزیه ماتریس، رسیدن به بهینه سراسری قابل تضمین نمی‌باشد، ولی با توجه به وجود اطلاعات نیمه‌نظارتی و اعمال فزون‌سازی انتظار می‌رود

حداقل به یک بهینه محلی خوب رسید. مشابه اکثر مقالات این حوزه، در این پایان نامه از روش حل گرادیان نزولی و ضرایب لاگرانژ، برای حل تابع هدف پیشنهادی (۳-۶) استفاده شده است [۵۹]. پس از به دست آوردن فرمول های به روزرسانی، حل این تجزیه به صورت یک الگوریتم تکراری صورت می پذیرد. به صورتی که، یکی از دو ماتریس مقداردهی اولیه می شود، سپس با ثابت در نظر گرفتن هر ماتریس، ماتریس دیگر با توجه به فرمول به دست آمده، به روزرسانی می شود [۴۲]. در این مسئله با مقداردهی اولیه تصادفی ماتریس H و سپس ثابت در نظر گرفتن H ، فرمول به روزرسانی ماتریس مراکز برچسب Z به صورت زیر می باشد:

$$Z = XH^T(HH^T)^{-1} \quad (۹-۳)$$

که HH^T یک ماتریس کوچک $l \times l$ ناتکین^۱ بوده که خاصیت وارون پذیری را دارد پس در محاسبه این ماتریس مشکلی وجود ندارد.

به دلیل شرط نامنفی بودن ماتریس H ، امکان محاسبه فرمول به روزرسانی این ماتریس از طریق شبه وارون وجود ندارد. با توجه به مسئله بهینه سازی محدود^۲، فرمول به روزرسانی زیر حاصل می شود:

$$H = H \odot \sqrt{\frac{[Z^T X]^+ + [Z^T Z]^- H + \lambda H W}{[Z^T X]^- + [Z^T Z]^+ H + \lambda H D + \varepsilon 1}} \quad (۱۰-۳)$$

که در این فرمول \odot بیانگر ضرب هادامارد^۳ یا ضرب درایه در درایه می باشد و ماتریس $1_{l \times n}$ ماتریس تمام یک و ε مقدار ثابت بسیار کوچک هستند. عبارت $\varepsilon 1$ به منظور جلوگیری که از تقسیم بر صفر شدن، به مخرج فرمول (۱۰-۳) اضافه شده اند. همچنین $[]^+$ نشان دهنده ماتریس با عناصر نامنفی و $[]^-$ نشان دهنده ماتریس با عنصر نامثبت می باشد و به صورت

$$B_{i,j}^+ = (|B_{i,j}| + B_{i,j})/2 \quad \text{و} \quad B_{i,j}^- = (|B_{i,j}| - B_{i,j})/2$$

محاسبه می شوند.

^۱ Nonsingular

^۲ Constrained optimization

^۳ Hadamard

Input: data matrix $X \in \mathbb{R}^{d \times n}$, label matrix $L \in \{0,1\}^{l \times n}$

Output: membership matrix H

Begin

Construct the affinity graph W by label matrix Y (2-3)

Initialize membership matrix H_0 randomly

Repeat

Reset certain labels by (7-3)

Update W by (6-3)

Update class matrix Z by (9-3)

Update membership matrix H by (10-3)

Until stopping criterion is reached

End

شرط خاتمه در این روش، به حداقل رسیدن خطای تجزیه $\|X - ZH\|_F^2 < \varepsilon$ و یا رسیدن به تعداد تکرار مشخص می‌باشد.

۳-۲-۶- مقداردهی اولیه مبتنی بر اطلاعات نظارتی

در اکثر مسائل تجزیه، برای مقداردهی اولیه ماتریس‌ها از روش مقداردهی تصادفی و یا از الگوریتم‌های خوشه‌بندی ساده مانند k-means استفاده می‌کنند. با توجه به استفاده از این تجزیه‌ها در کاربرد خوشه‌بندی و عدم وجود اطلاعات جانبی، این روش‌ها ساده‌ترین راه‌حل ممکن می‌باشند. حال با توجه به در دسترس بودن اطلاعات نظارتی موجود در روش پیشنهادی فعلی، استفاده از روش‌های تصادفی و خوشه‌بندی نمی‌تواند برای مسائل نیمه نظارتی راه‌حل بهینه‌ای باشند. در این روش، سعی شده است با وارد کردن اطلاعات نیمه نظارتی، ماتریس تعلقات اولیه را طوری مقداردهی کرد که شروع مناسبی برای مسئله مذکور باشد و موجب همگرایی سریع‌تر مسئله بهینه‌سازی شود. بدین منظور، از الگوریتم ساده و شناخته‌شده انتشار برچسب^۱ که یک روش مبتنی بر توزیع احتمال می‌باشد، استفاده شده است [۶۰]. این الگوریتم با استفاده از یک گراف وزن‌دار کامل، برچسب نمونه‌های برچسب‌دار را به نمونه‌های بدون برچسب انتشار می‌دهد و در نهایت تعلق هر نمونه به هر برچسب به صورت احتمالاتی حاصل می‌شود.

گام اول این الگوریتم ایجاد گراف یا ماتریس گذار^۲ می‌باشد. ابتدا شباهت هر جفت نمونه را با استفاده از تابع زیر به دست می‌آید:

$$d_{ij} = \exp - \|x_i - x_j\|^2 / \mu\sigma^2 \quad (11-3)$$

سپس برای نرمال‌سازی ماتریس D به صورت احتمالی، از فرمول زیر استفاده می‌شود:

¹ Label Propagation

² Transition matrix

$$p_{ij} = \frac{d_{ij}}{\sum_k d_{ik}} \quad (12-3)$$

که ماتریس گذار P نرمال شده و به صورت نامتقارن است و $\sum_k p_{ik} = 1$.
 گام دوم الگوریتم انتشار برچسب، منتشر کردن برچسب بر روی این گراف گذار می باشد. بدین منظور ماتریس برچسب $Y = [Y^{(nl)}, Y^{(nu)}] \in \mathbb{R}^{l \times n}$ ایجاد می شود که در آن $Y^{(nl)}$ ماتریس برچسب نمونه های برچسب دار و $Y^{(nu)}$ ماتریس برچسب نمونه های بدون چسب می باشد. Y_{ij} برابر 1 خواهد بود اگر نمونه i برچسب داشته باشد و در غیر این صورت، برابر 0 خواهد بود.
 اکنون با ضرب ماتریس برچسب Y در ماتریس گذار P به صورت زیر، برچسب نمونه های برچسب دار به نمونه های بدون برچسب انتشار پیدا می کند.

$$H_0 = YP \quad (13-3)$$

و این ماتریس حاصل می تواند به عنوان مقداردهی اولیه ماتریس تعلقات H مسئله بهینه سازی (3-6) مورد استفاده قرار گیرد.

نکته ۱: با توجه به کوچک بودن مقادیر نمونه های بدون برچسب ماتریس H_0 ، لازم است بر روی این ماتریس نرمال سازی احتمالاتی صورت گیرد تا تعلقات به صورت یکنواخت وارد مسئله شود.

$$h_{i,j}^{(0)} = \frac{h_{i,j}^{(0)}}{\sum_k h_{i,k}^{(0)}} \quad (14-3)$$

نکته ۲: الگوریتم انتشار برچسب فرایندی تکراری بوده اما در این پایان نامه - همان طور که در فرمول (3-12) مشاهده می شود - به دلیل پرهیز از هزینه محاسباتی اضافی، عمل انتشار برچسب فقط یک بار صورت می گیرد.

Input: data matrix $X \in \mathbb{R}^{d \times n}$, label matrix $L \in \{0,1\}^{l \times n}$

Output: membership matrix H

Begin

Construct the transition matrix P by (12-3)

Construct the affinity graph W by label matrix Y (2-3)

Initialize membership matrix H_0 (13-3)

Repeat

Reset certain labels by (7-3)

Update W by (6-3)

Update class matrix Z by (9-3)

Update membership matrix H by (10-3)

Until stopping criterion is reached

End

۳-۳- روش پیشنهادی دوم: تجزیه نیمه نظارتی مبتنی بر همبستگی محلی

در این بخش، الگوریتمی چندبرچسبه ارائه می شود که سعی می کند همبستگی محلی برچسب ها را استخراج نماید. در این مسئله فرض می شود هیچ اطلاعات خارجی از همبستگی محلی برچسب ها وجود ندارد و می توان نمونه ها را به دسته های مختلف تقسیم کرد که هر کدام از این دسته ها، همبستگی برچسب مشترکی دارند. روش پیشنهادی به گونه ای عمل می کند که هر دو عمل طبقه بندی نمونه ها و خوشه بندی دسته ها (استخراج همبستگی)، در یک مسئله تجزیه ماتریس واحد، به نام تجزیه عمیق نیمه-نامنفی ماتریس^۱ صورت پذیرد [۵۷].

۳-۳-۱- همبستگی محلی در تجزیه نیمه-نامنفی ماتریس

در بخش های قبل در مورد لزوم تعریف همبستگی در مسائل چندبرچسبه بحث شد، در این بخش سعی شده است نوعی همبستگی - به منظور بهبود دقت طبقه بندی - برای مسئله فعلی استخراج شود. در این مسئله، فرض می شود نمونه ها به m دسته خوشه بندی می شود به گونه ای که نمونه های داخل هر دسته زیرمجموعه ای از همبستگی های محلی را شامل می شود. با الهام گرفتن از [۱۱]، می توان برای یافتن این دسته ها از خوشه بندی تعلقات استفاده کرد. شناخته شده ترین خوشه بندی مبتنی بر بهینه سازی به صورت زیر تعریف می شود:

$$\min_{A,Q} \sum_{i=1}^n \sum_{j=1}^m q_{ji} \|h_i - a_j\|^2 \quad (15-3)$$

¹ Deep Semi-NMF

که در آن، مقدار q_{ji} معیاری است که میزان تأثیر زیرمجموعه همبستگی i بر نمونه x_i را نشان می‌دهد. h_i بردار تعلق نمونه i به برجسب‌هاست و a_j مرکز دسته (زیرمجموعه همبستگی) j ام می‌باشد. فرمول (۳-۱۵) مشابه الگوریتم مرسوم k -means می‌باشد ولی با این تفاوت که یک خوشه‌بندی نرم می‌باشد و بدین گونه تعریف می‌شود که اگر نمونه h_i با مرکز a_j فاصله کمی دارد پس بایستی تعلق نمونه q_{ji} ، مقدار بزرگی باشد و بالعکس.

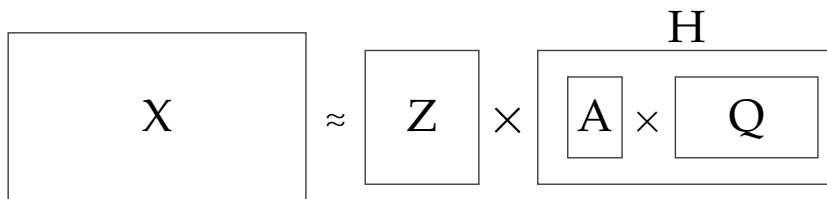
با استناد به مقاله [۴۲] می‌توان ادعا کرد که خوشه‌بندی (۳-۱۵)، معادل تجزیه نیمه-نامنفی ماتریس به صورت زیر می‌باشد:

$$\min_{A, Q} \sum_{i=1}^n \sum_{j=1}^m q_{ji} \|h_i - a_j\|^2 = \|H - AQ\|_F^2, \quad Q \geq 0 \quad (۱۶-۳)$$

در ادامه، به منظور طبقه‌بندی و استخراج همبستگی محلی به صورت هم‌زمان، از روش تجزیه عمیق نیمه-نامنفی ماتریس استفاده شده است:

$$\begin{aligned} X^{\pm} &\approx Z^{\pm} H^{+} \\ H^{+} &\approx A^{\pm} Q^{+} \end{aligned} \quad (۱۷-۳)$$

ابتدا به منظور طبقه‌بندی، ماتریس داده X به دو ماتریس مراکز برجسب Z و ماتریس "تعلقات نمونه‌ها به برجسب‌ها" یعنی H ، تجزیه می‌شود. سپس به منظور استخراج همبستگی محلی، ماتریس تعلقات H به دو ماتریس مراکز دسته A و ماتریس "تعلقات نمونه‌ها به دسته‌ها" یعنی Q ، تجزیه می‌شود.



شکل ۳-۲- ساختار سلسله مراتبی تجزیه

تابع هزینه اولیه، بدون در نظر گرفتن اطلاعات نظارتی به صورت زیر تعریف می‌شود:

$$\min_{Z, A, Q} \|X - ZAQ\|_F^2 + \frac{\lambda_1}{2} \|Q^T\|_{2,1}, \quad Q, [AQ] \geq 0 \quad (۱۸-۳)$$

محدودیت تعریف شده با توجه به نوع تجزیه، به صورت نامنفی بودن ماتریس $H=AQ$ و ماتریس Q می باشد.

در ادامه نرم $\|\bullet\|_{2,1}$ ماتریس استفاده شده است، که به صورت $\|B\|_{2,1} = \sum_i \|b^i\|_2$ محاسبه می شود و در بهینه سازی مسائل یادگیری ماشین مانند/تخاب ویژگی^۱ کاربرد فراوانی دارد [۶۱]. این نرم موجب تُنک شدن^۲ ماتریس به صورت ستونی می شود و در مسئله بهینه سازی فعلی، بر روی ترانهاده تعلقات دسته Q اعمال شده، که منجر به تُنک شدن سطری این ماتریس خواهد شد. این عمل باعث می شود تعلقات قوی هر نمونه حفظ شده و تعلقات ضعیف آن نمونه به حداقل مقدار خود برسد.

۳-۲-۳-۳ ایجاد گراف های برچسب احتمالاتی و خمینه

گراف برچسب احتمالاتی: در این روش مشابه روش اول، به منظور بازنمایی اطلاعات نیمه-نظارتی مسئله، از گراف برچسب استفاده خواهد شد. به منظور تأثیرپذیری بیشتر گراف و تناسب آن با مسئله چندبرچسبه، روشی احتمالاتی برای محاسبه شباهت مابین نمونه ها (در فضای برچسب)، به کار برده می شود [۶۲]. در این روش مشابه همبستگی ارائه شده در روش پیشنهادی اول، از ضرب ماتریس استفاده می شود $C = Y^T Y$. اینک با توجه به چندبرچسبه بودن نمونه ها و امکان این که دو نمونه بیش از یک برچسب مشترک داشته باشند، شباهت دو نمونه به صورت احتمالاتی و از طریق فرمول زیر محاسبه می شود:

$$w_{ij} = \frac{c_{ij}}{c_{ii} + c_{jj} - c_{ij}} \quad (19-3)$$

که در این فرمول c_{ii} نشان دهنده تعداد برچسب های نمونه i ، c_{jj} تعداد برچسب های نمونه j و c_{ij} تعداد برچسب های مشترک دو نمونه i و j می باشد [۶۲].

گراف خمینه: هرچند تجزیه ماتریس ابزاری مناسب برای یادگیری می باشد اما توانایی کشف ساختار تفکیک پذیر و هندسی ذاتی فضای داده ها را ندارد. یکی از کارآمدترین ابزارهای حفظ ساختار محلی هندسی داده، گراف k نزدیکترین همسایه^۳ می باشد که در این روش از گراف تُنک تر k نزدیکترین همسایه مشترک^۴ استفاده شده است [۶۳]. این گراف شامل n گره است که هر گره

¹ Feature selection

² Sparsity

³ k-nearest neighbors graph

⁴ k- mutual nearest neighbors graph

بیانگر یک نمونه می باشد. برای هر گره، k نزدیک ترین همسایه آن پیدا می شود و بین نمونه x_i و x_j یالی ایجاد می شود، اگر x_i در همسایگی x_j و x_j نیز در همسایگی x_i باشد. برای ایجاد این گراف از فرمول زیر استفاده شده است:

$$g_{ij} = \begin{cases} \exp\left(-\frac{\|x_i - x_j\|^2}{\sigma^2}\right) & x_j \in \text{knn}(x_i) \text{ and } x_i \in \text{knn}(x_j) \\ 0 & \text{otherwise} \end{cases} \quad (20-3)$$

که در این روش، وزن یال ایجاد شده از طریق کرنل حرارتی^۱ مقداردهی می شود و $\text{knn}(x_i)$ بیانگر مجموعه ی همسایگان نمونه x_i می باشد [۵۶].

۳-۳-۳- منظم سازی گراف ها

منظم سازی برچسب: مشابه منظم سازی روش اول، به منظور استفاده از اطلاعات نیمه نظارتی مسئله، یک جمله بهینه سازی برای جایگذاری اطلاعات برچسب نمونه های برچسب دار (که در قالب گراف برچسب احتمالاتی بازنمایی شده است) تعریف می شود.

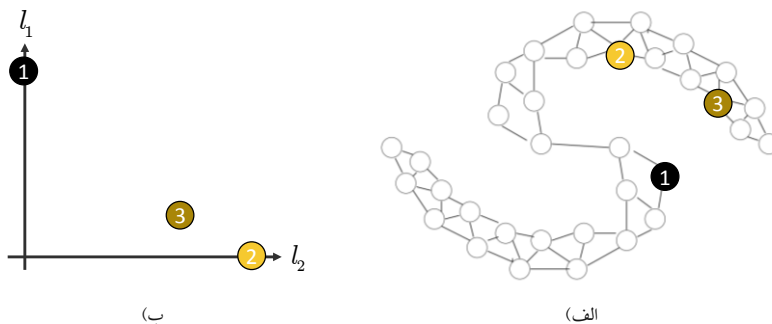
"اگر دو نمونه i و j در فضای برچسب، از معنایی مشابه هستند، پس این دو نمونه به میزان این شباهت، بایستی در دسته یا دسته های یکسانی قرار گرفته باشند" و به صورت زیر نوشته می شود:

$$\min_Q \frac{1}{2} \sum_{i,j=1}^n w_{i,j} \|q_i - q_j\|^2 = \text{Tr}(QLQ^T) \quad (21-3)$$

بدین صورت که، w_{ij} میزان شباهت معنایی نمونه i و j می باشد و q_i و q_j بردارهای تعلق نمونه های i و j به m دسته موجود هستند. مانند فرمول (۳-۵) از فرم ماتریسی این بهینه سازی استفاده می شود و بدین منظور، ماتریس درجه گراف برچسب D^W و گراف لاپلاسی L تعریف می شود که به صورت

$$L = D^W - W \quad \text{و} \quad d_{i,i}^W = \sum_{j=1}^n w_{i,j}$$

¹ Heat kernel



شکل ۳-۳- الف) مفهوم خمینه در فضای داده ب) مفهوم نمونه در فضای برچسب.

منظم‌سازی خمینه: همان‌طور که در بخش قبل، لزوم حفظ ساختار محلی-هندسی داده‌ها بیان شد، در این بخش نحوه جایگذاری اطلاعات خمینه در مسئله فعلی بیان می‌شود. این امر علاوه بر همواری مسئله بهینه‌سازی و منجر به مقاوم بودن مسئله طبقه‌بندی در برابر برچسب‌های نویزی [۴۷] نیز می‌شود. با استناد به منظم‌سازی خمینه ارائه‌شده در [۵۶]، فرمول جایگذاری اطلاعات ساختاری به‌صورت زیر تعریف می‌شود:

"اگر دو نمونه i و j بر روی خمینه، از نظر هندسی نزدیک هستند، پس این دو نمونه به میزان این فاصله، بایستی برچسب‌های مشابهی داشته باشند" و به‌صورت زیر فرمول‌سازی می‌شود:

$$\min_H \frac{1}{2} \sum_{i,j=1}^n g_{i,j} \|h_i - h_j\|^2 = \text{Tr}(\mathbf{H}\mathbf{S}\mathbf{H}^T) \quad (22-3)$$

بدین صورت که، $g_{i,j}$ میزان فاصله نمونه i و j می‌باشد و h_i و h_j بردارهای تعلق نمونه‌های i و j به l برچسب موجود هستند. مانند فرمول (۲۲-۳) از فرم ماتریسی این بهینه‌سازی استفاده می‌شود و بدین منظور، ماتریس درجه گراف خمینه D^G و گراف لاپلاسی S تعریف می‌شود که به‌صورت

$$S = D^G - G \text{ و } d_{i,i}^G = \sum_{j=1}^n g_{i,j} \text{ محاسبه می‌شوند.}$$

۳-۳-۴- تابع هدف پیشنهادی

اکنون با اضافه کردن جمله‌های منظم‌سازی گراف برچسب احتمالاتی و منظم‌سازی خمینه به تابع هزینه (۳-۱۸)، تابع هدف پیشنهادی زیر حاصل می‌شود که توانایی طبقه‌بندی داده‌های X به l برچسب را دارد و با خوشه‌بندی تعلقات حاصل به m دسته، همبستگی محلی را استخراج می‌کند. در ادامه، از نرم ۲ و به‌منظور قطعی شدن تعلقات نمونه‌ها به دسته‌ها بهره برده می‌شود. همچنین

به منظور حفظ ساختار کلاس ها و مقاوم بودن در مقابل برچسب های نویزی، منظم سازی خمینه اعمال می شود. در نهایت، از اطلاعات نیمه نظارتی مسئله برای ایجاد خوشه های دقیق تر استفاده شده است.

$$\min_{Z, H, A, Q} \|X - ZAQ\|_F^2 + \frac{\lambda_1}{2} \|Q^T\|_{2,1} + \frac{\lambda_2}{2} \text{Tr}(\text{HSH}^T) + \frac{\lambda_3}{2} \text{Tr}(\text{QLQ}^T) \quad Q, [QA] \geq 0 \quad (23-3)$$

محدودیت تعریف شده با توجه به نوع تجزیه، به صورت نامنفی بودن ماتریس Q و ماتریس H=AQ می باشد. پارامتر λ_1 تنظیم کننده میزان تُنکی سطری ماتریس Q و λ_2 و λ_3 به ترتیب پارامترهای کنترلی منظم سازی خمینه و منظم سازی گراف برچسب هستند.

۳-۵- فرمول های بروز رسانی

پس از تعریف تابع هدف نهایی، به دست آوردن فرمول های به روزرسانی اولین گام حل بهینه سازی می باشد. حل این تجزیه مشابه حل تجزیه عمیق نیمه-نامنفی [۵۷] می باشد و به صورت یک الگوریتم تکراری صورت می پذیرد. به صورتی که، دو ماتریس H و Q مقداردهی اولیه می شوند، سپس به صورت سلسله مراتبی با ثابت در نظر گرفتن دو ماتریس، ماتریس دیگر با توجه به فرمول به دست آمده، به روزرسانی می شود. در این بخش، با ثابت در نظر گرفتن ماتریس های H=AQ و Q مقادیر ماتریس Z به صورت زیر به روزرسانی خواهند شد.

ماتریس مراکز Z، مشابه ماتریس مراکز روش اول، با استفاده از شبه وارون به روزرسانی می شود:

$$Z = XQ^T A^T (AQQ^T A^T)^{-1} \quad (24-3)$$

فرمول به روزرسانی ماتریس تعلقات برچسب H همانند روش قبل و با شرط نامنفی بودن H به صورت زیر خواهد بود:

$$H = H \odot \sqrt{\frac{[Z^T X]^+ + [Z^T Z]^- H + \lambda_2 H G}{[Z^T X]^- + [Z^T Z]^+ H + \lambda_2 H D^G}} \quad (25-3)$$

مشابه فرمول به روزرسانی مراکز برچسب Z، به روزرسانی مراکز دسته A نیز به همان روش بدون محدودیت انجام می شود، با این تفاوت که باید از دو شبه وارون ماتریس استفاده شود:

$$A = Z^+ X Q^+ \quad (26-3)$$

به منظور به روزرسانی تعلقات دسته Q به دلیل محدودیت نامنفی بودن این ماتریس مشابه ماتریس H به روزرسانی انجام می شود و با توجه به منظم سازی های مبتنی بر Q، فرمول آن به شکل زیر خواهد بود:

$$Q = Q \odot \sqrt{\frac{[A^T Z^T X]^+ + [A^T Z^T Z A]^- Q + \lambda_3 Q W}{[A^T Z^T X]^- + [A^T Z^T Z A]^+ Q + \lambda_1 V Q + \lambda_3 Q D^W}} \quad (27-3)$$

که $V \in \mathbb{R}^{l \times l}$ یک ماتریس قطری است به صورتی که $v_{ii} = 1 / 2 \|q_i^T\|_2$ و q_i^T سطر نام از ماتریس Q می باشد. هرچند به صورت تجربی احتمال صفر شدن $\|q_i^T\|_2$ وجود ندارد، اما از لحاظ تئوری این امکان وجود دارد و برای جلوگیری از تقسیم بر صفر شدن عبارت فوق، ثابت بسیار کوچک ε به آن اضافه می شود و از عبارت $v_{ii} = 1 / 2 \|q_i^T\|_2 + \varepsilon$ برای محاسبه ماتریس قطری V به کار برده شده در (3-25) استفاده می شود.

الگوریتم ۳. شبه کد روش پیشنهادی دوم

Input: data matrix $X \in \mathbb{R}^{d \times n}$, label matrix $L \in \{0,1\}^{l \times n}$

Output: category centers A, category membership Q

Begin

Construct semantic relevance graph W by (19-3)

Construct nearest neighbor graph G by (20-3)

Construct transition matrix P by (12-3)

Initialize label membership H_0 by (13-3)

Initialize category membership Q_0 by k-means

Repeat

Update Z by (24-3)

Update H by (25-3)

Update A by (26-3)

Update Q by (27-3)

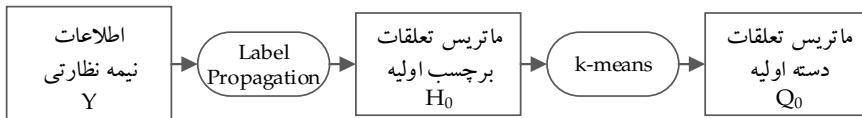
Until Stopping criterion is reached

End

۳-۶- مقداردهی اولیه مبتنی بر اطلاعات نظارتی و خوشه بندی

همان طور که در روش اول بیان شد، با توجه به در دسترس بودن اطلاعات نظارتی موجود در مسئله فعلی، استفاده از روش های تصادفی و خوشه بندی نمی تواند برای مسائل نیمه نظارتی راه حل بهینه ای باشند. در این روش نیز، سعی شده است با وارد کردن اطلاعات نیمه نظارتی، ماتریس تعلقات برچسب اولیه H را طوری مقداردهی کرد که شروع مناسبی برای مسئله مذکور باشد. اما برای

مقداردهی اولیه ماتریس تعلقات دسته اولیه، هیچ دانش قبلی وجود ندارد در این روش مانند روش [۴۲] برای مقداردهی اولیه این ماتریس از خوشه‌بندی شناخته‌شده k-mean استفاده شده است.



شکل ۳-۴- نمودار مقداردهی اولیه ماتریس‌های H و Q

پس مشابه روش اول، با استفاده از اطلاعات نیمه‌نظارتی و اجرای الگوریتم انتشار برچسب، ماتریس تعلقات برچسب اولیه H_0 مقداردهی اولیه می‌شود. سپس با اجرای الگوریتم k-means بر روی ماتریس H_0 ، ماتریس Q_0 حاصل می‌شود. بدین صورت که، با مشخص شدن تعلق نمونه i به خوشه j ، مقدار $q_{ji}=1$ خواهد بود. به منظور جلوگیری از بایاس شدن خوشه‌بندی، مقدار ثابت 0.2 به کلیه عناصر ماتریس Q_0 اضافه می‌شود.

فصل چهارم: نتایج و تفسیر آن‌ها

۴-۱- مقدمه

در این فصل، عملکرد روش‌های پیشنهادی مورد ارزیابی قرار می‌گیرد. روش پیشنهادی اول با عنوان DWSF با یک الگوریتم نوین نیمه‌نظارتی مقایسه می‌شود. روش پیشنهادی دوم نیز با یک روش نظارتی چندبرچسبه و دو روش نیمه‌نظارتی چندبرچسبه مورد مقایسه قرار می‌گیرد. الگوریتم‌ها در متلب ۲۰۱۷ پیاده‌سازی کرده‌ایم. آزمایشات روی کامپیوتری با مشخصات Intel® Core i5 CPU 2.5 GHz و ۶ گیگابایت حافظه‌ی RAM، انجام شده‌اند. برای ارزیابی روش‌های ارائه شده، مجموعه داده‌های چندکلاسه و چندبرچسبه به کار برده می‌شوند. مشخصات مجموعه‌های داده‌ای مورد استفاده، معیارهای ارزیابی و نتایج نیز در ادامه توصیف می‌شوند.

۴-۲- مجموعه داده‌ها

۴-۲-۱- مجموعه داده‌های چندکلاسه

به منظور ارزیابی روش پیشنهادی اول، آزمایشاتی بر روی چهار مجموعه داده انجام شده است که از مخزن داده دانشگاه کالیفرنیا (ایرواین) [۶۴] تهیه شده‌اند. این مجموعه داده‌ها حاوی اطلاعات کاربردی واقعی در زمینه‌های مختلف هستند و به طور گسترده‌ای برای آزمودن عملکرد الگوریتم‌های مختلف یادگیری ماشین استفاده می‌شوند. جدول ۴-۱ مجموعه داده‌های انتخاب شده با تعداد نمونه، تعداد ویژگی و تعداد کلاس را نشان می‌دهد. در این انتخاب سعی شده است از داده‌های چندکلاسه نیز استفاده شود.

مجموعه داده قلب (Heart) شامل اطلاعات وجود یا عدم وجود بیماری قلبی است. مجموعه داده‌ها شامل مطالعاتی است که بر روی ۲۷۰ بیمار انجام شده است. موارد مورد بررسی شامل ۱۳ ویژگی از جمله سن، جنس و چندین علائم حیاتی بیماری قلبی است.

مجموعه داده یون کره (Ionosphere) توسط یک سیستم در آزمایشگاه Goose Bay جمع‌آوری شده است که شامل یک آرایه فازی از ۱۶ آنتن با فرکانس بالاست. اهداف الکترون‌های آزاد در یونوسفر بودند. بازدهی آنتن "بد" است اگر سیگنال‌های مستقیماً از یون کره عبور می‌کنند و بازدهی "خوب" است اگر شواهدی نشان دهد اطلاعات از طریق یونوسفر بازتاب شده‌است.

مجموعه داده بذر (Seeds) شامل سه نوع مختلف گندم کاما، رزا و کانادایی است و ۷۰ عدد از هر کدام به صورت تصادفی برای آزمایش انتخاب شده است. آشکارسازی ساختار داخلی هسته با استفاده از تکنیک اشعه ایکس ملایم صورت گرفته است.

مجموعه داده شیشه (Glass) به منظور طبقه‌بندی شیشه تهیه شده است. نتایج تجزیه شیمیایی خرده شیشه‌ها (به عنوان هشت ارزش واقعی عناصر تشکیل دهنده) و شاخص انکسار، برای طبقه‌بندی نمونه به شیشه‌های شناور یا غیرشناور ساختمانی، شیشه‌های خودرو، ظروف شیشه‌ای، لوازم آشپزخانه، یا لامپ استفاده شده‌است.

جدول ۴-۱- مجموعه داده‌های استفاده شده در روش پیشنهادی اول

| نام مجموعه داده | کاربرد | تعداد نمونه | تعداد ویژگی | تعداد کلاس |
|-----------------|------------|-------------|-------------|------------|
| Heart | پزشکی | ۲۷۰ | ۱۳ | ۲ |
| Ionosphere | مخابرات | ۳۵۱ | ۳۵ | ۲ |
| Seeds | زیست‌شناسی | ۲۱۰ | ۷ | ۳ |
| Glass | علوم قضائی | ۲۱۴ | ۹ | ۶ |

۴-۲-۲- مجموعه داده‌های چندبرچسبه

به منظور ارزیابی روش پیشنهادی دوم، آزمایشاتی بر روی دو مجموعه داده انجام شده. این مجموعه داده‌ها حاوی اطلاعات کاربردی واقعی در زمینه‌های مختلف هستند و به طور گسترده‌ای برای آزمودن عملکرد الگوریتم‌های مختلف یادگیری چندبرچسبه استفاده می‌شوند. جدول ۴-۲ مجموعه داده‌های انتخاب شده با تعداد نمونه، تعداد ویژگی و تعداد برچسب را نشان می‌دهد. در این انتخاب سعی شده است از داده‌های مقیاس بزرگ با تعداد برچسب زیاد استفاده شود.

Corel5k [65] مجموعه داده‌ای چندبرچسبه شامل ۵۰۰۰ تصویر است که در کل ۳۷۴ برچسب وجود دارد و هر تصویر ۴ تا ۵ کلمه کلیدی دارد. تصاویر با استفاده از برش‌های نرمال شده تقسیم شده‌اند و فقط نواحی که بیش از یک مقدار آستانه بوده‌اند، استفاده شده‌اند. معمولاً برای هر تصویر ۵-۱۰ ناحیه وجود دارد و نواحی که برای توصیف تصویر نهایی استفاده می‌شوند، با استفاده از خوشه‌بندی به ۴۹۹ جاب تقسیم شده است.

Enron [66] مجموعه داده‌ای چندبرچسبه شامل ۱۷۰۲ ایمیل است که از پایگاه داده Enron، به روش شبه‌انگیزی انتخاب شده است. این انتخاب با تمرکز بر ایمیل‌های مربوط به کسب‌وکار و بحران انرژی کالیفرنیا و به وسیله دانشجویان دوره پردازش زبان طبیعی دانشگاه برکلی صورت گرفته است. پیام‌های انتخاب شده به وسیله دو دانشجو با برچسب‌های تعریف شده، دسته‌بندی شده است.

جدول ۴-۲- مجموعه داده‌های استفاده شده در روش پیشنهادی دوم

| نام مجموعه داده | کاربرد | تعداد نمونه | تعداد ویژگی | تعداد کلاس |
|-----------------|-------------------|-------------|-------------|------------|
| Corel5K | برچسب‌گذاری تصویر | ۵۰۰۰ | ۴۹۹ | ۳۷۴ |
| Enron | طبقه‌بندی متن | ۱۷۰۲ | ۱۰۰۱ | ۵۳ |

۳-۴- معیارهای ارزیابی

جهت ارزیابی روش‌های پیشنهادی با روش‌های دیگر موجود، از معیار شناخته شده F-measure که در حوزه‌ی طبقه‌بندی چندکلاسه کاربرد فراوانی دارد، استفاده شده است. همچنین از معیارهای میانگین درستی و خطای رتبه‌بندی برای ارزیابی مسائل چندبرچسب استفاده شده است.

۱-۳-۴- معیارهای ارزیابی مسائل چندکلاسه

معیار F-measure یک معیار مبتنی بر بازیابی اطلاعات است به این معنی که به تعیین شباهت بین نتایج حاصل از روش‌ها و اطلاعات واقعی می‌پردازد [۶۷]؛ بنابراین نتایج حاصل از الگوریتم‌ها در مقایسه با اطلاعات واقعی در یکی از چهار حالت زیر قرار می‌گیرند:

- T_P : تعداد نمونه‌های که توسط الگوریتم به درستی برچسب یکسان گرفته‌اند.
- T_N : تعداد نمونه‌هایی که در واقعیت برچسب متفاوت داشته و توسط الگوریتم نیز برچسب‌های متفاوتی گرفته‌اند
- F_P : تعداد نمونه‌هایی که در واقعیت برچسب کلاس‌های متفاوتی دارند، درحالی‌که توسط الگوریتم برچسب کلاس یکسان گرفته‌اند.
- F_N : تعداد نمونه‌هایی که در واقعیت برچسب کلاس‌های یکسان دارند درحالی‌که توسط الگوریتم برچسب کلاس‌های متفاوت گرفته‌اند.

F-Measure، پارامتر مناسبی برای ارزیابی کیفیت کلاس‌بندی می‌باشد و همچنین توصیف‌کننده میانگین وزن‌دار مابین دو کمیت Precision و Recall است. که با توجه به رابطه زیر محاسبه می‌شود:

$$F - measure = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (۱-۴)$$

برای یک الگوریتم کلاس‌بندی‌کننده در شرایط ایده‌آل، مقدار این کمیت برابر با ۱ و در بدترین وضعیت برابر با صفر می‌باشد.

Precision: نسبت تعداد نمونه‌های درست برچسب‌گذاری شده توسط الگوریتم از یک کلاس به تعداد کل نمونه‌هایی است که چه به صورت اشتباه و چه به صورت صحیح توسط الگوریتم برچسب آن کلاس را خورده‌اند.

$$Precision = \frac{T_p}{T_p + F_p} \quad (۲-۴)$$

Recall: نسبت تعداد نمونه‌های درست برچسب‌گذاری شده توسط الگوریتم از یک کلاس به تعداد کل نمونه‌هایی است که در واقعیت در آن کلاس قرار دارند. که به صورت زیر محاسبه می‌گردد:

$$Recall = \frac{T_p}{T_p + F_n} \quad (۳-۴)$$

۴-۳-۲- معیارهای ارزیابی مسائل چندبرچسبه

در این بخش دو معیار شناخته شده مبتنی بر تأثیرپذیری برچسب معرفی می‌شود که در بسیاری از روش‌های طبقه‌بندی چندبرچسبه به کار برده شده‌اند [۶۸]. برای آشنایی با این معیارها، یک پیش-بینی کننده چند برچسبه F را فرض کنید، در این مسئله Y_i نشان‌دهنده i امین بردار سطری ماتریس برچسب می‌باشد. Y_i^+ مشخص کننده ی مجموعه اندیس برچسب‌های مرتبط Y_i و Y_i^- مشخص کننده ی مجموعه اندیس برچسب‌های نامرتب Y_i هستند. به عبارت دیگر $Y_i^+ = \{j | y_{ij} = 1\}$ و $Y_i^- = \{j | y_{ij} = 0\}$. بیانگر کاردینالیته یک مجموعه است، بنابراین تعداد برچسب‌های مرتبط x_i ، $|Y_i^+|$ می‌باشد.

از دیدگاه تأثیرپذیری برچسب، برای هر جفت $(u, v) \in Y_i^+ \times Y_i^-$ داریم $f_u(x_i) > f_v(x_i)$. بنابراین مجموعه معکوس S_{rank}^i تهی است و کاردینالیته مجموعه صفر می‌باشد، بدین معنی است که مجموع کاردینالیته همه مجموعه معکوس‌ها $rloss(F) = 0$ ؛ و خطای رتبه‌بندی بر روی مجموعه داده بهینه‌سازی شده است.

$$rloss(F) = \frac{1}{m} \sum_{i=1}^m \frac{|S_{rank}^i|}{|Y_i^+| |Y_i^-|} \quad (۴-۴)$$

$$S_{rank}^i = \{(u, v) | f_u(x_i) \leq f_v(x_i), (u, v) \in Y_i^+ \times Y_i^-\}$$

پس با توجه به مطالب گفته شده، هدف طبقه‌بند چندبرچسبه کم کردن خطای رتبه‌بندی می‌باشد.

برای درک بهتر این معیار، فرض کنید Z برچسب مرتبط نمونه i است و معادله آن به شکل زیر است:

$$rank_F(x_i, j) = \left| \{k \in Y_i^+ \mid rank_F(x_i, k) \leq rank_F(x_i, j)\} \right| \quad (5-4)$$

از این رو $rank_F(x_i, j)$ دقیقاً تعریف $S_{precision}^{ij}$ می باشد و $avgprec(F) = 1$ ، یعنی میانگین درستی بر روی مجموعه داده بهینه سازی شده است

$$avgprec(F) = \frac{1}{m} \sum_{i=1}^m \frac{1}{|Y_i^+|} \sum_{j \in Y_i^+} \frac{|S_{precision}^{ij}|}{rank_F(x_i, j)} \quad (6-4)$$

$$S_{precision}^{ij} = \{k \in Y_i^+ \mid rank_F(x_i, k) \leq rank_F(x_i, j)\}$$

پس با توجه به مطالب گفته شده، هدف طبقه بند چندبرچسبه بیشینه کردن میانگین درستی می باشد.

۴-۴-۴ نتایج

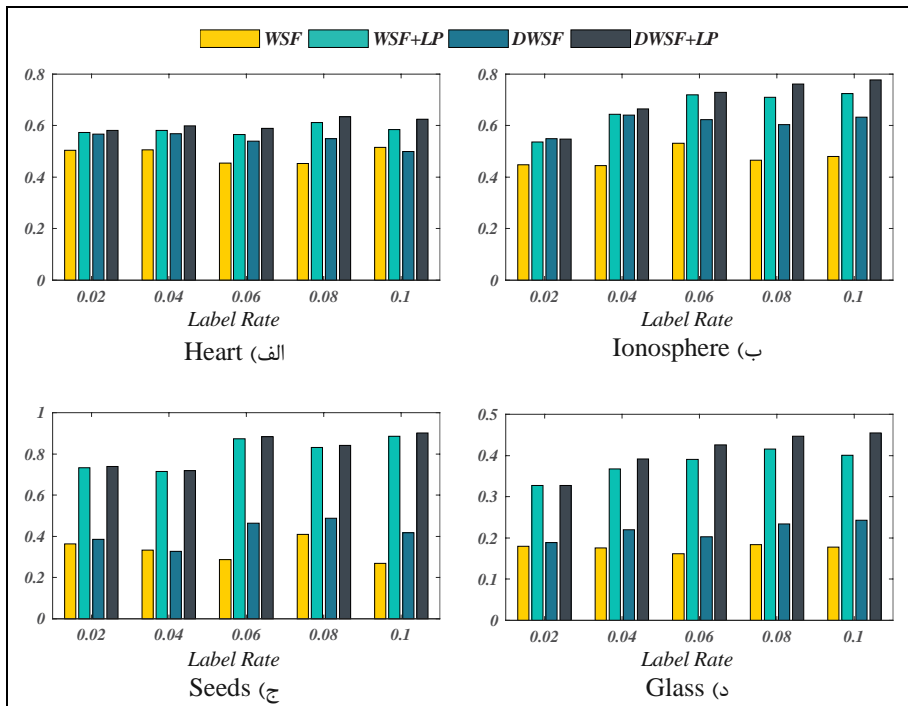
در این قسمت، نتایج آزمایشات روی هر دو روش پیشنهادی به صورت مجزا ارائه می گردد.

۴-۴-۱-۱ نتایج روش پیشنهادی اول

در این بخش، روش پیشنهادی اول با عنوان DWSF، بر روی مجموعه داده های قلب، یون کره، بذر و شیشه اجرا شده و نتایج حاصل با روش WSF که در سال ۲۰۱۷ منتشر شده [۵۷]، مقایسه شده است. به منظور بررسی میزان تأثیر همبستگی پویا و همچنین راهکار مقداردهی اولیه ماتریس ها، روش پیشنهادی به صورت سه الگوریتم مجزا تعریف شده است. در الگوریتم اول (DWSF) فقط با استفاده از همبستگی پویا طبقه بندی را انجام می شود، در الگوریتم دوم (WSF+LP) فقط تأثیر مقداردهی اولیه ماتریس ها بررسی می شود و در الگوریتم سوم (DWSF+LP) از ترکیب همبستگی پویا و مقداردهی اولیه استفاده شده است. در این آزمایش به صورت تصادفی درصد متفاوتی از داده های برچسب دار با نرخ برچسب ۲٪ تا ۱۰٪ نمونه برداری شده و برای هر نرخ برچسب ثابت، ۱۰ آزمایش مستقل با نمونه برداری تصادفی اجرا شده است. متوسط F-measure و متوسط مدت زمان اجرای هر الگوریتم، به ترتیب در شکل های ۴-۱ و ۴-۲ گزارش شده است.

همانطور که در شکل ۴-۱ مشاهده می شود، الگوریتم های فوق الذکر بر روی چهار مجموعه داده اجرا شده است. هر دو عمل مقداردهی اولیه و استخراج همبستگی در همه مجموعه داده ها عملکرد بهتری دارند. مقداردهی اولیه به دلیل استفاده از اطلاعات موجود از داده های برچسب دار به جای

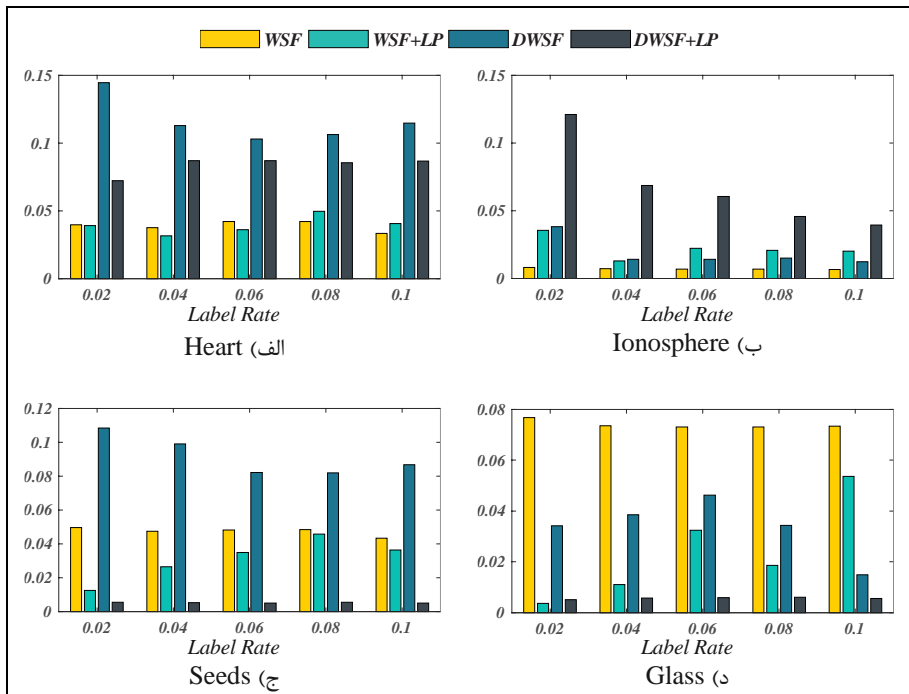
استفاده از مقداردهی تصادفی، یک شروع کور نیست و می‌تواند موجب بهبود عملکرد شود. همچنین استفاده از برجسب‌های قطعی و غیرقطعی در محاسبه همبستگی پویا، موجب افزایش تأثیرگذاری فرایند منظم‌سازی می‌شود و همانطور که مشخص است، بر روی دو مجموعه داده شیشه و بذر که چندکلاسه هستند، نتایج بهتری حاصل شده که نشان دهنده توانایی روش پیشنهادی اول در حل مسائل چندبرچسبه موثر می‌باشد.



شکل ۴-۱- ارزیابی روش پیشنهادی اول بر روی مجموعه داده‌ها براساس معیار ارزیابی F-measure

در شکل ۴-۲ الگوریتم‌های معرفی شده از لحاظ مدت زمان اجرا با هم مقایسه شده‌اند. به دلیل استفاده از فرایند استخراج همبستگی در روش پیشنهادی اول، این روش نسبت روش WSF بار محاسباتی سرباری دارد و همانطور که در مجموعه‌های دو کلاسه قلب و یون‌کره مشاهده می‌شود موجب افزایش مدت زمان اجرا می‌شود. اما در مجموعه داده‌های چندکلاسه بذر و شیشه، همگرایی سریع‌تر صورت می‌پذیرد و این موجب کاهش مدت زمان اجرا شده است.

پس با توجه به نتایج حاصل می‌توان نتیجه گرفت روش پیشنهادی بر روی مجموعه داده‌های چندکلاسه، دقت قابل توجهی دارد و همچنین مدت زمان اجرا بر روی این داده‌ها بسیار کمتر است. پس می‌توان در مسائل چندبرچسبه نیز از این راهکار استفاده کرد.

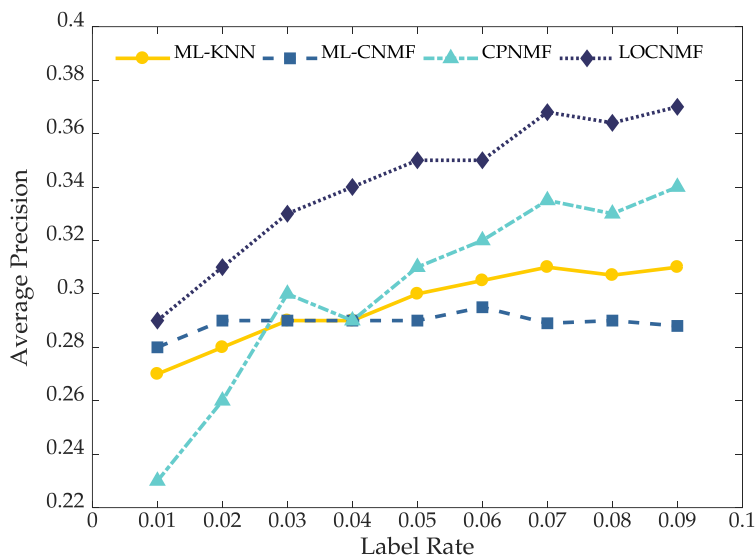


شکل ۴-۲- ارزیابی روش پیشنهادی اول بر روی مجموعه داده‌ها براساس مدت زمان اجرا

۴-۴-۲- نتایج روش پیشنهادی دوم

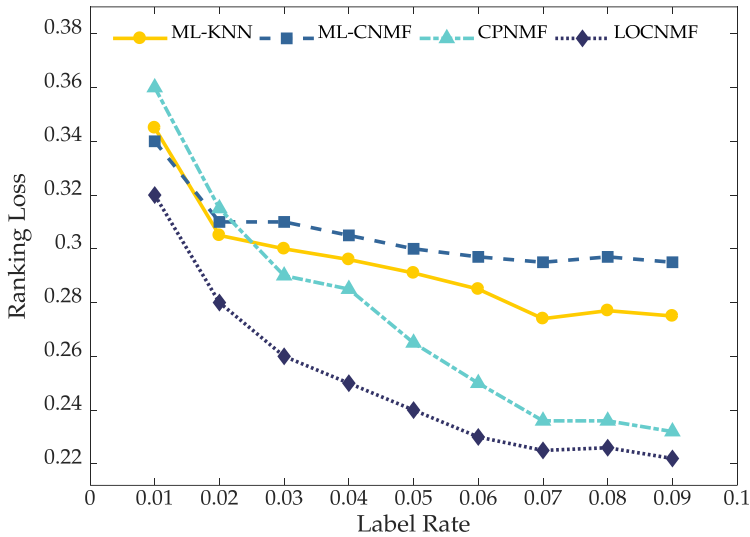
در این بخش، روش پیشنهادی دوم با عنوان LOCNMF، بر روی مجموعه داده‌های Corel5K و Enron اجرا شده و نتایج حاصل با روش‌های ML-KNN [۴۴]، ML-NMF [۴۶] و CPNMF [۴۷] مقایسه شده است. روش اول (ML-KNN) شناخته‌ترین روش یادگیری چندبرچسبه است. روش دوم (ML-NMF) یادگیری نیمه‌نظارتی چندبرچسبه مبتنی بر تجزیه است، که از همبستگی سراسری استفاده می‌کند و روش سوم (CPNMF) روشی نوین بوده و توسعه یافته روش قبلی می‌باشد. در این آزمایش به صورت تصادفی درصد متفاوتی از داده‌های برچسب‌دار با نرخ برچسب ۱٪ تا ۹٪ نمونه‌برداری شده و برای هر نرخ برچسب ثابت، ۱۰ آزمایش مستقل با نمونه‌برداری تصادفی اجرا شده

است. متوسط میانگین درستی و متوسط خطای رتبه‌بندی هر روش، در شکل های ۳-۴، ۴-۴، ۴-۵ و ۶-۴ گزارش شده است.



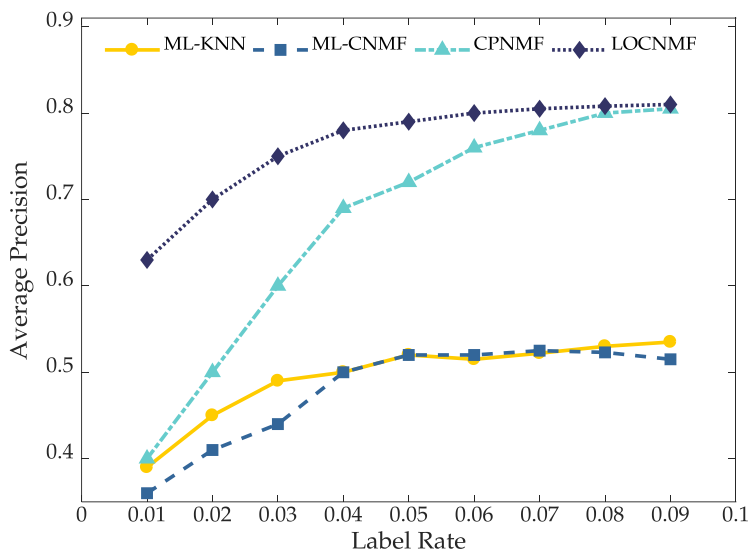
شکل ۳-۴- ارزیابی روش پیشنهادی دوم بر روی مجموعه داده Corel5K براساس معیار میانگین درستی

همانطور که در شکل ۳-۴ مشاهده می‌شود، روش‌های فوق‌الذکر بر روی مجموعه داده Corel5K اجرا شده است. روش پیشنهادی با هر نرخ برجستگی، نسبت به سایر روش‌ها دقت بالاتر و عملکرد بهتری دارد. همچنین در شکل ۴-۴ مشخص است که این روش بر روی مجموعه داده، خطای رتبه‌بندی کمتری دارد.

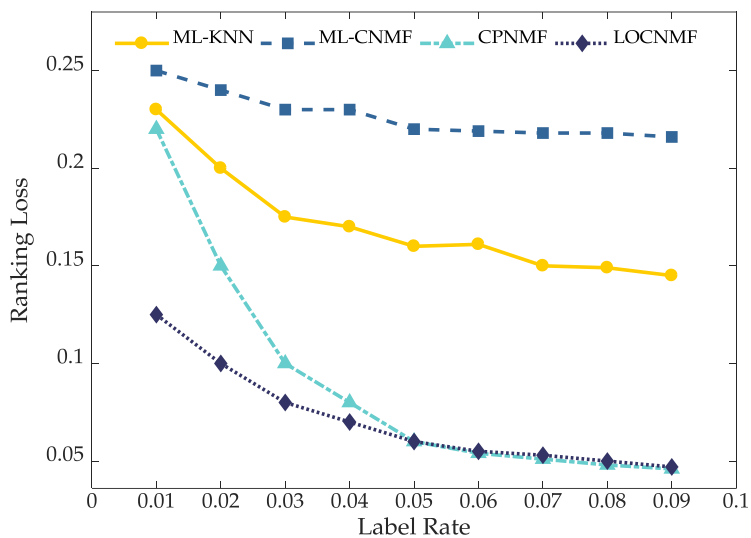


شکل ۴-۴- ارزیابی روش پیشنهادی دوم بر روی مجموعه داده Corel5K براساس معیار خطای رتبه‌بندی

همچنین در شکل ۴-۵ مشاهده می‌شود، روش‌های پیشنهادی دوم بر روی مجموعه داده Enron اجرا شده است و باز نسبت به سایر روش‌ها دقت بالاتر و عملکرد بهتری دارد. یک مشاهده قابل توجه این است که وقتی نرخ برچسب کم است، روش پیشنهادی می‌تواند بهبود قابل توجهی نسبت به روش‌های موجود داشته باشد. یکی از دلایل این است که LOCNMF هم از ساختار محلی و هم همبستگی محلی استفاده می‌کند که در روش‌های نیمه نظارتی و چندبرچسبه، ضروری است. همچنین در شکل ۴-۶ مشخص است که این روش بر روی مجموعه داده، خطای رتبه‌بندی کمتری دارد.



شکل ۴-۵- ارزیابی روش پیشنهادی دوم بر روی مجموعه داده Enron براساس معیار میانگین درستی



شکل ۴-۶- ارزیابی روش پیشنهادی دوم بر روی مجموعه داده Enron براساس معیار خطای رتبه‌بندی

فصل پنجم: نتیجه‌گیری و پیشنهادات

۵-۱- نتیجه‌گیری

در مسائل نیمه‌نظارتی چندبرچسبه، کار با تعداد محدود نمونه‌های برچسب‌دار نسب به تعداد نمونه‌های بدون برچسب و همچنین تعداد بالای برچسب‌ها و نحوه استخراج همبستگی مابین برچسب‌ها چالش‌های این گونه یادگیری‌ها است. در این پژوهش دو روش یادگیری نیمه‌نظارتی ارائه شد که از تکنیک تجزیه نیمه‌نامنفی ماتریس بهره گرفته‌اند.

در روش اول سعی شد با الهام گرفتن از روش‌های فزون‌سازی، روشی مبتنی بر تکرار ارائه شود، که از تعلق حاصل از هر تکرار به عنوان اطلاعات غیرقطعی استفاده شود و همبستگی به صورت پویا استخراج شود. سپس اطلاعات همبستگی دوباره به الگوریتم یادگیری داده می‌شود. در ادامه به منظور سرعت بخشیدن به فرایند تجزیه مسئله، از راهکاری مبتنی بر اطلاعات نیمه‌نظارتی بجای مقداردهی تصادفی معمول استفاده شد که موجب بهبود عملکرد طبقه‌بندی نیز می‌شود. روش پیشنهادی اول به صورت تحلیلی با روشی نوین مقایسه شد و نتایج حاصل نشان داد که این روش پیشنهادی بر روی مجموعه داده‌های چندکلاسه کارایی بالا دارد و در مدت زمان مناسبی به همگرایی می‌رسد. همچنین این روش به دلیل استفاده از همبستگی پویا، توانایی طبقه‌بندی مجموعه داده‌های چندبرچسبه را خواهد داشت.

روش پیشنهادی دوم، به منظور بهبود الگوریتم‌های یادگیری نیمه‌نظارتی چندبرچسبه ارائه شد. در روش‌های مبتنی بر تجزیه فعلی، از همبستگی سراسری مابین برچسب‌ها برای طبقه‌بندی کارآمد مسائل چندبرچسبه استفاده شده است. در این روش به دلیل حساس بودن همبستگی سراسری در مقابل برچسب‌های نویزی، از مفهوم همبستگی محلی استفاده شد، که نوعی همبستگی‌های مرتبه بالا

می‌باشد. به منظور حفظ یکپارچگی مسئله، فرایندهای استخراج همبستگی محلی و طبقه‌بندی، در یک چارچوب واحد با عنوان تجزیه نیمه نامنفی عمیق انجام می‌پذیرد. همچنین در این روش به منظور حفظ ساختار فضای داده و مقاوم بودن مسئله در برابر برچسب‌های نویزی، از منظم‌سازی خمینه استفاده شده است. این روش پیشنهادی بر روی مجموعه داده‌های چندبرچسبه در زمینه‌های حاشیه‌نویسی تصویر و طبقه‌بندی متن اجرا شده و با روش‌های شناخته شده و نوین این حوزه مقایسه شد. آزمایشات نشان می‌دهد روش پیشنهادی دوم با وجود تعداد کم نمونه‌های برچسب‌دار، کارایی مناسبی دارد.

۵-۲- پیشنهادات برای تحقیقات آتی

روش‌های پیشنهادی این پایان‌نامه می‌توانند در زمینه‌ی یادگیری نیمه‌نظارتی و یادگیری چندبرچسبه، مؤثر واقع شوند. با این وجود، راه برای ارائه‌ی روش‌های بهتر در این حوزه، هنوز باز است. در این قسمت پیشنهاداتی برای ارائه‌ی روش‌های با کارایی بالاتر بیان می‌شود.

- در روش پیشنهادی اول، در تکرارهای اولیه مکان نوین دار بودن تعلقات وجود دارد. می‌توان برای افزایش دقت همبستگی پویا، به مرور زمان اثربخشی همبستگی را افزایش داد.
- به منظور استخراج اطلاعات همبستگی دقیق‌تر و بهبود فرایند برچسب‌گذاری، می‌توان از همبستگی محلی و سراسری به صورت هم‌زمان استفاده کرد.
- با استفاده از تکنیک‌های کاهش پیچیدگی و موازی‌سازی و همچنین استفاده از تجهیزات محاسباتی، می‌توان روش فعلی را با مجموعه داده‌های حوزه کلان داده سازگار کرد.
- می‌توان از مفاهیم نوین یادگیری، از قبیل یادگیری خودآموز یا خودمختار برای بهبود دقت طبقه‌بندی استفاده کرد.

- [1] K. Dembczynski, W. Cheng, and E. Hüllermeier, "Bayes Optimal Multilabel Classification via Probabilistic Classifier Chains," in International Conference on Machine Learning (ICML), 2010, vol. 10, pp. 279-286.
- [2] G. Tsoumakas, A. Dimou, E. Spyromitros, V. Mezaris, I. Kompatsiaris, and I. Vlahavas, "Correlation-based pruning of stacked binary relevance models for multi-label learning," in International Workshop on Learning from Multi-Label Data, 2009, pp. 101-116.
- [3] M.-L. Zhang and K. Zhang, "Multi-label learning by exploiting label dependency," in International Conference on Knowledge Discovery and Data Mining (SIGKDD) 2010, pp. 999-1008.
- [4] S. Zhu, X. Ji, W. Xu, and Y. Gong, "Multi-labelled classification using maximum entropy method," in Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, 2005, pp. 274-281.
- [5] R. Yan, J. Tesic, and J. R. Smith, "Model-shared subspace boosting for multi-label classification," in International Conference on Knowledge Discovery and Data Mining (SIGKDD) 2007, pp. 834-843.
- [6] G.-J. Qi, X.-S. Hua, Y. Rui, J. Tang, T. Mei, and H.-J. Zhang, "Correlative multi-label video annotation," in International Conference on Multimedia, 2007, pp. 17-26.
- [7] S.-J. Yang, Y. Jiang, and Z.-H. Zhou, "Multi-instance multi-label learning with weak label," in International Joint Conference on Artificial Intelligence (IJCAI), 2013, pp. 1862-1868.
- [8] X. Li and Y. Guo, "Active Learning with Multi-Label SVM Classification," in International Joint Conference on Artificial Intelligence (IJCAI), 2013, pp. 1479-1485.
- [9] S. Ji, L. Sun, R. Jin, and J. Ye, "Multi-label multiple kernel learning," in Advances in Neural Information Processing Systems (NIPS), 2009, pp. 777-784.
- [10] B. Hariharan, L. Zelnik-Manor, M. Varma, and S. Vishwanathan, "Large scale max-margin multi-label classification with priors," in International Conference on Machine Learning (ICML), 2010, pp. 423-430.
- [11] S.-J. Huang, Z.-H. Zhou, and Z. Zhou, "Multi-Label Learning by Exploiting Label Correlations Locally," in AAAI Conference on Artificial Intelligence, 2012, pp. 949-955.
- [12] Y.-N. Chen and H.-T. Lin, "Feature-aware label space dimension reduction for multi-label classification," in Advances in Neural Information Processing Systems (NIPS), 2012, pp. 1529-1537.
- [13] Z. Lin, G. Ding, M. Hu, and J. Wang, "Multi-label classification via feature-aware implicit label space encoding," in International Conference on Machine Learning (ICML), 2014, pp. 325-333.
- [14] Y. Zhang and J. Schneider, "Multi-label output codes using canonical correlation analysis," in International Conference on Artificial Intelligence and Statistics, 2011, pp. 873-882.

- [15] O. Chapelle, B. Schölkopf, and A. Zien, *Semi-Supervised Learning*. MIT Press, 2010, p. 528.
- [16] X. Zhu, Z. Ghahramani, and J. D. Lafferty, "Semi-supervised learning using gaussian fields and harmonic functions," in *International Conference on Machine Learning (ICML)*, 2003, pp. 912-919.
- [17] X. Zhou and M. Belkin, "Semi-supervised learning," in *Academic Press Library in Signal Processing*, vol. 1, 2014, pp. 1239-1269.
- [18] 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.
- [19] D. Donoho and V. Stodden, "When does non-negative matrix factorization give a correct decomposition into parts?," in *Advances in Neural Information Processing Systems (NIPS)*, 2004, pp. 1141-1148.
- [20] Y.-X. Wang and Y.-J. Zhang, "Nonnegative matrix factorization: A comprehensive review," *IEEE Transactions on Knowledge and Data Engineering*, vol. 25, no. 6, pp. 1336-1353, 2013.
- [21] B. Klingenberg, J. Curry, and A. Dougherty, "Non-negative matrix factorization: Ill-posedness and a geometric algorithm," *Pattern Recognition*, vol. 42, no. 5, pp. 918-928, 2009.
- [22] N. Vasiloglou, A. G. Gray, and D. V. Anderson, "Non-negative matrix factorization, convexity and isometry," in *International Conference on Data Mining (ICDM)*, 2009, pp. 673-684.
- [23] W. Xu, X. Liu, and Y. Gong, "Document clustering based on non-negative matrix factorization," in *Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2003, pp. 267-273.
- [24] I. Buciú, "Non-negative matrix factorization, a new tool for feature extraction: theory and applications," *International Journal of Computers, Communications and Control*, vol. 3, no. 3, pp. 67-74, 2008.
- [25] H. Liu, Z. Wu, X. Li, D. Cai, and T. S. Huang, "Constrained nonnegative matrix factorization for image representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 7, pp. 1299-1311, 2012.
- [26] P. O. Hoyer, "Non-negative sparse coding," in *Workshop on Neural Networks for Signal Processing*, 2002, pp. 557-565.
- [27] N. Mohammadiha and A. Leijon, "Nonnegative matrix factorization using projected gradient algorithms with sparseness constraints," in *International Symposium on Signal Processing and Information Technology (ISSPIT)*, 2009, pp. 418-423.
- [28] S. Z. Li, X. W. Hou, H. J. Zhang, and Q. S. Cheng, "Learning spatially localized, parts-based representation," in *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2001, pp. 207-212.
- [29] C. Ding, T. Li, W. Peng, and H. Park, "Orthogonal nonnegative matrix t-factorizations for clustering," in *International Conference on Knowledge Discovery and Data Mining (SIGKDD)*, 2006, pp. 126-135.
- [30] I. Kotsia, S. Zafeiriou, and I. Pitas, "A novel discriminant non-negative matrix factorization algorithm with applications to facial image characterization problems,"

- IEEE Transactions on Information Forensics and Security, vol. 2, no. 3, pp. 588-595, 2007.
- [31] Y. W. Y. Jia and C. H. M. Turk, "Fisher non-negative matrix factorization for learning local features," in Asian Conference on Computer Vision (ACCV), 2004, pp. 27-30.
 - [32] D. Cai, X. He, X. Wu, and J. Han, "Non-negative matrix factorization on manifold," in International Conference on Data Mining (ICDM) 2008, pp. 63-72.
 - [33] S. Zhang, W. Wang, J. Ford, and F. Makedon, "Learning from incomplete ratings using non-negative matrix factorization," in International Conference on Data Mining (ICDM), 2006, pp. 549-553.
 - [34] Y.-D. Kim and S. Choi, "Weighted nonnegative matrix factorization," in International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2009, pp. 1541-1544.
 - [35] P. Smaragdis, "Non-negative matrix factor deconvolution; extraction of multiple sound sources from monophonic inputs," in International Conference on Independent Component Analysis and Signal Separation, 2004, pp.499-494 .
 - [36] J. Yoo and S. Choi, "Orthogonal nonnegative matrix tri-factorization for co-clustering: Multiplicative updates on stiefel manifolds," Information processing & management, vol. 46, no. 5, pp. 559-570, 2010.
 - [37] T. Hazan, S. Polak, and A. Shashua, "Sparse image coding using a 3D non-negative tensor factorization," in International Conference on Computer Vision (ICCV) 2005, vol. 1, pp. 50-57.
 - [38] A. Shashua and T. Hazan, "Non-negative tensor factorization with applications to statistics and computer vision," in International Conference on Machine learning (ICML), 2005, pp. 792-799.
 - [39] L. Li and Y.-J. Zhang, "Non-negative matrix-set factorization," in International Conference on Image and Graphics (ICIG), 2007, pp. 564-569.
 - [40] D. Zhang, Z.-H. Zhou, and S. Chen, "Non-negative matrix factorization on kernels," in Pacific Rim International Conference on Artificial Intelligence, 2006, pp. 404-412.
 - [41] I. Buciu, N. Nikolaidis, and I. Pitas, "Nonnegative matrix factorization in polynomial feature space," IEEE Transactions on Neural Networks, vol. 19, no. 6, pp. 1090-1100, 2008.
 - [42] C. H. Q. Ding, T. Li, and M. I. Jordan, "Convex and Semi-Nonnegative Matrix Factorizations," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 32 ,no. 1, pp. 45-55, 2010.
 - [43] M. R. Boutell, J. Luo, X. Shen, and C. M. Brown, "Learning multi-label scene classification," Pattern Recognition, vol. 37, no. 9, pp. 1757-1771, 2004.
 - [44] 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.
 - [45] J. Fürnkranz, E. Hüllermeier, E. L. Mencía, and K. Brinker, "Multilabel classification via calibrated label ranking," Machine Learning, vol. 73, no. 2, pp. 133-153, 2008.
 - [46] Y. Liu, R. Jin, and L. Yang, "Semi-supervised multi-label learning by constrained non-negative matrix factorization," in AAAI Conference on Artificial Intelligence, 2006, vol. 6, pp. 421-426.

- [47] X. Zhang, N. Guan, Z. Luo, and X. Yang, "Constrained Projective Non-negative Matrix Factorization for Semi-supervised Multi-label Learning," in International Conference on Machine Learning and Applications (ICMLA), 2015, pp. 588-593.
- [48] W. Cheng and E. Hüllermeier, "Combining instance-based learning and logistic regression for multilabel classification," *Machine Learning*, vol. 76, no. 2-3, pp. 211-225, 2009.
- [49] S. Ji, L. Tang, S. Yu, and J. Ye, "Extracting shared subspace for multi-label classification," in International Conference on Knowledge Discovery and Data Mining (SIGKDD) 2008, pp. 381-389.
- [50] E. L. Allwein, R. E. Schapire, and Y. Singer, "Reducing multiclass to binary: A unifying approach for margin classifiers," *Journal of Machine Learning Research*, vol. 1, no. Dec, pp. 113-141, 2000.
- [51] 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.
- [52] X. Kong, M. K. Ng, and Z.-H. Zhou, "Transductive multilabel learning via label set propagation," *IEEE Transactions on Knowledge and Data Engineering*, vol. 25, no. 3, pp. 704-719, 2013.
- [53] K. Crammer and Y. Singer, "A new family of online algorithms for category ranking," in Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, 2002, pp. 151-158.
- [54] J. Rousu, C. Saunders, S. Szedmak, and J. Shawe-Taylor, "Kernel-based learning of hierarchical multilabel classification models," *Journal of Machine Learning Research*, vol. 7, no. Jul, pp. 1601-1626, 2006.
- [55] F. Kang, R. Jin, and R. Sukthankar, "Correlated label propagation with application to multi-label learning," in Conference on Computer Vision and Pattern Recognition (CVPR), 2006, vol. 2, pp. 1719-1726.
- [56] 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.
- [57] G. Trigeorgis, K. Bousmalis, S. Zafeiriou, and B. W. Schuller, "A Deep Matrix Factorization Method for Learning Attribute Representations," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 3, pp. 417-429, 2017.
- [58] B. Wang, Z. Tu, and J. K. Tsotsos, "Dynamic Label Propagation for Semi-supervised Multi-class Multi-label Classification," in International Conference on Computer Vision (ICCV), 2013, pp. 425-432.
- [59] S. Sra, S. Nowozin, and S. J. Wright, *Optimization for machine learning*. MIT Press, 2012, p. 512.
- [60] X. Zhu, J. Lafferty, and R. Rosenfeld, "Semi-supervised learning with graphs," Carnegie Mellon University, language technologies institute, school of computer science, 2005.
- [61] F. Nie, H. Huang, X. Cai, and C. H. Ding, "Efficient and robust feature selection via joint ℓ_2 , ℓ_1 -norms minimization," in Advances in Neural Information Processing Systems (NIPS), 2010, pp. 1813-1821.

- [62] Z. Li and J. Tang, "Weakly Supervised Deep Matrix Factorization for Social Image Understanding," *IEEE Transactions on Image Processing*, vol. 26, no. 1, pp. 276-288, 2017.
- [63] L. Berton and A. de Andrade Lopes, "Graph construction based on labeled instances for semi-supervised learning," in *International Conference on Pattern Recognition (ICPR)*, 2014, pp. 2477-2482.
- [64] D. Dua and E. Karra Taniskidou, "UCI Machine Learning Repository," S. o. I. a. C. S. University of California, Ed., ed, 2017.
- [65] P. Duygulu, K. Barnard, J. F. de Freitas, and D. A. Forsyth, "Object recognition as machine translation: Learning a lexicon for a fixed image vocabulary," in *European Conference on Computer Vision (ECCV)*, 2002, pp. 97-112.
- [66] B. Klimt and Y. Yang, "The enron corpus: A new dataset for email classification research," *European Conference on Machine Learning (ECML)*, pp. 217-226, 2004.
- [67] D. M. W. Powers, "Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation," *International Journal of Machine Learning Technology*, vol. 2, no. 1, pp. 37-63, 2011.
- [68] X.-Z. Wu and Z.-H. Zhou, "A Unified View of Multi-Label Performance Measures," in *International Conference on Machine Learning (ICML)*, 2017, pp. 3780-3788.

پیوست: واژه نامه‌ی انگلیسی به فارسی

| واژگان | معادل فارسی |
|-----------------------------------|-------------------------------|
| Active learning | یادگیری فعال |
| Boosting | فزون‌سازی |
| Classification | طبقه‌بندی |
| Classifier | طبقه‌بند |
| Clustering | خوشه‌بندی |
| Collaborative filtering | فیلترسازی مشارکتی |
| Constrained optimization | بهینه‌سازی مقید |
| Convex | محدب |
| Correlation | همبستگی |
| Deep Semi-NMF | تجزیه نیمه-نامنفی عمیق ماتریس |
| Dynamic correlation | همبستگی پویا |
| Error-correcting coding | کد گذاری تصحیح خطا |
| Feature selection | انتخاب ویژگی |
| Feedback | بازخورد |
| Frobenius norm | نرم فروبنیوس |
| Gene expression analysis | تحلیل بیان ژن |
| Global correlation | همبستگی سراسری |
| Gradient descent | گرادیان نزولی |
| Graph regularization | منظم‌سازی گراف |
| Hadamard product | ضرب هادامارد |
| Heat kernel | هسته حرارتی |
| Hyperparameter | ابر پارامتر |
| Image Annotation | حاشیه‌نویسی تصویر |
| Inductive learning | یادگیری استقرایی |
| In-sample | نمونه مشاهده شده |
| k- mutual nearest neighbors graph | گراف k نزدیکترین همسایه مشترک |

| | |
|----------------------------------|------------------------------|
| k-nearest neighbors graph | گراف k نزدیکترین همسایه |
| Label Propagation | انتشار برچسب |
| Local correlation | همبستگی محلی |
| Machine learning | یادگیری ماشین |
| Manifold | خمینه |
| Maximum A posteriori Principle | اصل بیشینه گر احتمال پسین |
| Multi-label learning | یادگیری چندبرچسبه |
| Nonnegative matrix factorization | تجزیه نامنفی ماتریس |
| Nonsingular | ناتکین |
| NP-hardness | سختی زمان چندجمله‌ای غیرقطعی |
| One-versus-one | یک در مقابل یک |
| One-versus-the-rest | یک در مقابل بقیه |
| Out-of-sample | نمونه مشاهده نشده |
| Part-of-speech tagging | برچسب‌زنی اجزای کلام |
| Principal component analysis | تجزیه و تحلیل مؤلفه اصلی |
| Protein interaction | تعامل پروتئین |
| Semi-supervised learning | یادگیری نیمه‌نظارتی |
| Singular value decomposition | تجزیه مقدار منفرد |
| Smoothness | همواری |
| Source separation | جداسازی منبع |
| Sparse coding | کدگذاری تُنک |
| Sparsity | تُنکی |
| Structured data | داده ساختاری |
| Supervised learning | یادگیری نظارتی |
| Text Classification | طبقه‌بندی متن |
| Transductive learning | یادگیری انتقالی |
| Transition matrix | ماتریس گذار |
| Unsupervised learning | یادگیری بدون نظارت |
| Weighted low-rank approximation | تقریب کم‌رتبه وزن‌دار |

بسمه تعالی

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

(لازم است به عنوان صفحه اول پروپوزال و به عنوان چهارمین برگ پایان نامه و پس از صفحه مشخصات پایان نامه بوده و به دقت

مطالعه و امضا شود)

اینجانب **سید اکبر بیگلری** دانشجوی مقطع **کارشناسی ارشد رشته حقوق کیفری** متعهد میشوم:

- ۱- صداقت، امانتداری و بی طرفی را در انجام پژوهش و انتشار نتایج حاصل از آن رعایت نمایم.
- ۲- در نگارش نتیجه پژوهش های حاصل از موضوع پایان نامه، از باز نویسی نوشته های دیگران بدون ذکر منبع، بازی با الفاظ، زیاده نویسی، کلی گویی و جزم اندیشی و تصرف گرایي پرهیز نمایم و نتایج پژوهشی خود را در موعد مقرر و با اطلاع استاد راهنما منتشر نمایم.
- ۳- تمامی یافته های 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)، راهنما و استاد (استادان) مشاور را در لیست مولفین مقاله ذکر نمایم و از آوردن اسامی افرادی که نقش مؤثری در انجام پژوهش نداشته اند، جداً خودداری نمایم.
- ۵- در بخش سیاستگذاری مقاله، از تمامی افراد و سازمانهایی که در اجرای پژوهش مساعدتی میدول داشته اند یا ذکر نوع مشارکت تشکر و قدر دانی نمایم.
- ۶- از انتشار همپوشان یا ارسال همزمان یک مقاله بد چند مجله ویا 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.



اعضاء و اثر انگشت دانشجو

دستورالعمل نحوه برخورد با موارد تخلفی دانشجویان تحصیلات تکمیلی در هنگام انتشار نتایج پژوهش

- ۱- در موارد زیر دانشگاه کردستان با مجله مربوطه مکاتبه و درخواست خارج نمونه مقاله را نموده و موضوع را به محل کار یا تحصیل بعدی دانشجو اطلاع خواهد داد.
الف- چاپ مقاله بدون اطلاع و تأیید استادان راهنما،
ب- چاپ نتایج حاصل از پژوهشی های 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 تصمیم گیری خواهد نمود.



دانشگاه کردستان
دانشکده مهندسی
گروه مهندسی نرم افزار کامپیوتر

پایان نامه کارشناسی ارشد رشته مهندسی کامپیوتر گرایش هوش مصنوعی

عنوان:

یک راهکار یادگیری نیمه نظارتی مبتنی بر گراف
برای طبقه بندی چند برچسبه

پژوهشگر:

سید امجد سیدی

استاد راهنما:

دکتر پرهام مرادی

استاد مشاور:

دکتر فردین اخلاقیان طاب

بهمن ۱۳۹۶



دانشگاه کردستان
دانشکده مهندسی
گروه مهندسی نرم افزار کامپیوتر

پایان نامه کارشناسی ارشد رشته مهندسی کامپیوتر گرایش هوش مصنوعی




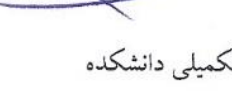
عنوان:

یک راهکار یادگیری نیمه نظارتی مبتنی بر گراف
برای مسائل چندبرچسبه

پژوهشگر:

سیدامجدسیدی

در تاریخ ۱۳۹۶/۱۱/۲۸ توسط کمیته تخصصی و هیئت داوران زیر مورد بررسی قرار گرفت و با
درجه عالی به تصویب رسید.

| امضاء | مرتبه علمی | نام و نام خانوادگی | هیئت داوران |
|---|------------|--------------------------|---------------------|
|  | دانشیار | دکتر پرهام مرادی | ۱- استاد راهنما |
|  | استادیار | دکتر فردین اخلاقیان طاب | ۲- استاد مشاور |
|  | استادیار | دکتر شاهرخ اسماعیلی | ۳- استاد داور خارجی |
|  | استادیار | دکتر علیرضا عبدالله پوری | ۴- استاد داور داخلی |

مهر و امضاء معاون آموزشی و تحصیلات تکمیلی دانشکده

دکتر مهرداد خاموشی

مهر و امضاء گروه

دکتر علیرضا عبدالله پوری



دانشگاه کردستان
دانشکده مهندسی
گروه مهندسی نرم افزار کامپیوتر

پایان نامه کارشناسی ارشد رشته مهندسی کامپیوتر گرایش هوش مصنوعی

عنوان

یک راهکار یادگیری نیمه نظارتی مبتنی بر گراف
برای طبقه بندی چندبرچسبه

پژوهشگر

سیدامجدسیدی

استاد راهنما

دکتر پرهام مرادی

استاد مشاور

دکتر فردین اخلاقیان طاب

بهمن ۱۳۹۶