



دانشگاه صنعتی شریف
دانشکده مهندسی کامپیوتر
سمینار کارشناسی ارشد گرایش هوش مصنوعی

عنوان:

بر چسب گذاری داده با استفاده از یادگیری نیمه نظارتی مبتنی بر منیفلد در داده‌های چندطیفی سنجش از راه دور
Data Labelling Using Manifold-Based Semi-Supervised Learning in Multispectral Remote Sensing

نگارش:

احمد خواجه‌نژاد
۸۹۲۱۰۸۴۱

استاد راهنما:

دکتر محمدعلی صفری - دکتر حمیدرضا ربیعی

استاد ممتحن داخلی:

دکتر مهدی جلیلی

چکیده: مساله‌ی تشخیص برچسب هر نقطه از یک تصویر سنجش از دور که در ابتدا فقط برچسب برخی از نقاط آن را می‌دانیم، با نام دسته‌بندی تصاویر سنجش از دور شناخته شده است. به دلیل ویژگی‌های این مساله، مثل تعداد زیاد داده‌ها و تعداد کم داده‌های برچسب‌دار، روش‌های یادگیری نیمه‌نظارتی در آن کاربرد زیادی دارند. به طور کلی، روش‌های نیمه‌نظارتی استفاده شده در این مساله را می‌توان به سه دسته‌ی روش‌های مولد، روش‌های مبتنی بر جداسازی در ناحیه‌ی کم تراکم و روش‌های مبتنی بر گراف (منیفلد) تقسیم کرد. در این گزارش، ابتدا به تعریف مساله و ویژگی‌های خاص آن پرداخته، سپس در مورد سه دسته روش مذکور توضیحاتی ارائه کرده و با تمرکز بر روش‌های مبتنی بر گراف، به مقایسه‌ی تئوری و عملی این روش‌ها می‌پردازیم.

واژه‌های کلیدی: سنجش از دور، یادگیری نیمه‌نظارتی، منیفلد، چندطیفی، ابرطیفی

۱ مقدمه

به طور کلی سنجش از دور، به معنی کسب اطلاعات از طریق تحلیل داده‌های به دست آمده از تصاویر و امواج برداشت شده از راه دور از شیء، مکان یا پدیده‌ای می‌باشد. در این پژوهش، هدف تحلیل داده‌های به دست آمده از عکس‌های هوایی یا ماهواره‌ای است. در عکس‌های RGB معمولی، فقط میزان نور بازتاب شده در محدوده‌ی طول موج‌های قابل دید، در سه بازه‌ی مربوط به رنگ‌های قرمز، سبز و آبی ذخیره شده است. اما در عکس‌های مورد مطالعه‌ی ما میزان نور بازتاب شده از هر نقطه در باندهای طیفی (یعنی همان بازه‌های طول موج) مرئی و غیر مرئی مختلف ذخیره شده است، بنابراین اگر مثلاً اطلاعات مربوط به ۸ بازه‌ی فرکانسی ذخیره شده باشد، هر پیکسل به صورت یک بردار به طول ۸ از اعداد می‌باشد. به همین دلیل این عکس‌ها را چندطیفی^۱ می‌گویند. به بیان دیگر می‌توان این عکس‌ها را به صورت تعدادی عکس (سیاه و سفید) از یک منطقه در نظر گرفت که هر عکس مربوط به یک باند طیفی است. معمولاً اگر باندهای طیفی عریض بوده و تعداد آن‌ها کم (در حدود ۱۰ باند) باشد، تصویر را چندطیفی گویند، اما اگر باندها باریک بوده و تعداد آن‌ها زیاد باشد، تصویر را ابرطیفی^۲ می‌گویند. زمانی که کاربرد مورد نظر ما به صورتی است که نیاز به بررسی تغییرات کوچک در باندهای باریک دارد، استفاده از تصاویر ابرطیفی لازم می‌شود. تمرکز ما هم در این پژوهش بر تحلیل داده‌های ابرطیفی است.

در تحلیل عکس‌های ابرطیفی، چند هدف وجود دارد که مساله‌ی اصلی مورد بررسی ما در این پروژه، دسته‌بندی^۳ می‌باشد. این مساله را برچسب‌گذاری^۴ یا نوعی قطعه‌بندی^۵ هم می‌توان نامید. در این مساله هدف این است که از روی تصاویر موجود، برچسب مربوط به هر نقطه (پیکسل) را تشخیص دهیم. بسته به کاربردهای مختلف، این برچسب می‌تواند بافت زمین، پوشش گیاهی، عمق آب یا موارد دیگر باشد. البته در مواردی مثل تشخیص عمق آب، مجموعه‌ی مقادیر مجاز به عنوان برچسب یک پیکسل، یک بازه‌ی پیوسته می‌باشد. بنابراین در این حالت با یک مساله از نوع رگرسیون مواجه هستیم، نه دسته‌بندی. در دسته‌بندی، هر نقطه برچسبی دارد که یکی از چند مقدار مشخص را می‌تواند داشته باشد.

به عنوان داده‌های ورودی مساله، ما مقادیر طیفی مربوط به همه‌ی پیکسل‌ها و برچسب بعضی از پیکسل‌ها را می‌دانیم، و هدف این است که برچسب نقاط بدون برچسب را پیش‌بینی کنیم. البته در اغلب کاربردهای دسته‌بندی، می‌توانیم مساله را به این صورت بیان کنیم که تصویر به نواحی‌ای تقسیم می‌شود و نقاط هر ناحیه، برچسب یکسانی دارند (نقاط دو ناحیه‌ی جدا از هم می‌توانند برچسب یکسانی داشته باشند) و ما می‌خواهیم با داشتن نمایش طیفی تمام نقاط و برچسب برخی از آن‌ها، تمام نواحی موجود و برچسب مربوط به هر ناحیه را پیدا کنیم.

زیاد بودن تعداد ابعاد فضای داده‌ها در مسائل یادگیری ماشین، پیچیدگی مسائل را زیاد می‌کند. این مفهوم شناخته شده‌ای است که از آن با نام مشکل بعد^۶ یاد می‌شود. بنابراین پیدا کردن تعداد کمی پایه‌ی مناسب که بتوان داده‌ها را در فضای ساخته شده از آن‌ها بیان کرد، به طوری که پراکندگی داده‌ها حفظ شود، از اهمیت زیادی برخوردار است. این کار را استخراج ویژگی^۷ می‌نامند. در مسائل مربوط به سنجش از دور، چون با داده‌های ابرطیفی سر و کار داریم، مساله‌ی استخراج ویژگی اهمیت زیادی پیدا می‌کند. بنابراین مساله‌ی استخراج ویژگی نیز مورد توجه ما خواهد بود.

در سنجش از دور، مسائل دیگری نیز وجود دارد که البته ما به آن‌ها نخواهیم پرداخت، اما با اخذ از [۱] بعضی از آن‌ها را بیان می‌کنیم. مثلاً در مساله‌ی تشخیص تغییر^۸، عکس‌هایی از یک منطقه در زمان‌های متفاوت در دست داریم و هدف این است که تغییرات نواحی را در طول زمان تشخیص دهیم. برای مثال در تشخیص گسترش مناطق شهری یا تغییر نوع پوشش گیاهی یا تغییر نواحی در اثر زلزله با این مساله مواجه هستیم.

از دیگر مسائل موجود در این زمینه خوشه‌بندی^۹ و تجسم داده^{۱۰} است. خوشه‌بندی ممکن است به صورت کاربرد مستقل یا پیش‌پردازشی برای کاربردهای دیگر مثل دسته‌بندی انجام شود. همچنین تجسم داده‌ها می‌تواند برای تحلیل بصری داده‌ها مفید باشد.

در یک عکس، هر پیکسل ترکیبی از الگوی بازتابی مواد تشکیل دهنده آن است. در مساله‌ی تجزیه‌ی سیگنالی^{۱۱} هدف آن است که پیکسل‌های خالصی را به عنوان پایه پیدا کنند که بتوانند نمایش طیفی هر پیکسل را با استفاده از ترکیبی خطی از آن‌ها نمایش دهند. این پیکسل‌های پایه را به اصطلاح عنصر نهایی^{۱۲} می‌گویند.

همچنین در مسائلی که مدلی برای بازتاب طیفی نقاط فرض کرده‌اند و می‌خواهند پارامترهای مدل را پیدا کنند، مساله‌ی رگرسیون و معکوس‌یابی مدل^{۱۳} مطرح می‌شود.

همان‌طور که گفته شد، تاکید ما در این پروژه بر مساله‌ی دسته‌بندی برای تصاویر ابرطیفی است، که در نتیجه با استخراج ویژگی هم درگیر خواهیم بود. رویکرد ما به این مساله با مفروضات و رویکردهای خاصی است که در ادامه به آن‌ها اشاره می‌شود.

۲ ویژگی‌های مساله

مساله‌ی مورد بررسی، در نگاه اول یک مساله‌ی دسته‌بندی ساده می‌باشد، اما در واقع به دلیل ویژگی‌هایی که دارد، مساله‌ای خاص می‌باشد. این ویژگی‌ها بدین صورت می‌باشند:

- زیاد بودن تعداد داده‌ها: چون در این مساله، هدف برچسب گذاری پیکسل‌های یک تصویر می‌باشد، معمولا با تعداد زیادی داده مواجه می‌باشیم. زیاد بودن تعداد داده‌ها، باعث افزایش بیش از حد زمان اجرا و حافظه‌ی مورد استفاده در بسیاری از الگوریتم‌ها می‌شود. به همین دلیل، یکی از چالش‌های مساله، و در نتیجه یک زمینه‌ی تلاش پژوهشگران در این مساله، ارائه‌ی روش‌های تقریبی برای الگوریتم‌های مختلف، با زمان اجرا و حافظه‌ی کمتر می‌باشد.
- کم بودن تعداد داده‌های برچسب‌دار آموزش: در یک کاربرد واقعی، برچسب داده‌هایی که قرار است در فاز آموزش استفاده شوند، یا به وسیله‌ی بررسی عکس توسط متخصص و تعیین برچسب برخی نقاط تعیین می‌شود، یا از طریق بررسی نقاط واقعی در مکان (مثلا در کاربرد تشخیص منابع زیرزمینی) یا به صورت ترکیب این دو روش. در هر صورت، تعیین برچسب نقاط کاری سخت و هزینه‌بر است. بنابراین به دنبال یافتن روشی هستیم که بتواند با استفاده از تعداد کمی داده‌ی برچسب‌دار دسته بندی را به خوبی انجام دهد.
- بعد بالای داده‌ها: زیاد بودن تعداد ابعاد داده‌ها هم دقت یادگیری را کاهش می‌دهد، و هم برای ذخیره و ارسال داده‌ها از ماهواره به زمین ایجاد مشکل میکند. به همین دلیل برخی تلاش‌ها صرفا در جهت کاهش بعد داده‌ها به نحوی که جدایی پذیری آن‌ها حفظ شود، می‌باشد.
- وابستگی مکانی نقاط به یکدیگر: در ساده‌ترین حالت، می‌توان بردار ویژگی مربوط به هر پیکسل را همان نمایش طیفی آن (مقدار نور بازتاب شده از آن نقطه در هر بازه‌ی طول موج) در نظر گرفت. اما چون نقاط تصویر قابل تقسیم به نواحی‌ای می‌باشد که در هر ناحیه برچسب تمام نقاط یکسان است، پس در واقع احتمال متفاوت بودن برچسب دو نقطه که فاصله‌ی مکانی کمی دارند، زیاد نیست. از این ویژگی می‌توان در یادگیری استفاده کرد. از ایده‌های اصلی که برای استفاده از این ویژگی مطرح شده است، می‌توان ناحیه‌بندی^{۱۴} و میانگین‌گیری را نام برد. در ناحیه‌بندی، ابتدا در یک فاز غیرنظارتی^{۱۵}، تصویر به نواحی مختلفی افراز می‌شود، سپس از هر ناحیه نمایندگانی انتخاب می‌شوند و یادگیری فقط با توجه به آن‌ها انجام می‌شود.
- روش میانگین‌گیری به این صورت است که بردار ویژگی مربوط به هر پیکسل، از ترکیب نمایش طیفی همان پیکسل، با میانگین نمایش طیفی نقاط در یک همسایگی آن نقطه ساخته می‌شود. در روش‌های یادگیری مبتنی بر هسته، می‌توان یک هسته مبتنی بر نمایش طیفی نقاط، و یک هسته مبتنی بر میانگین نمایش طیفی در همسایگی هر نقطه تعریف کرده و این دو هسته را به روش‌های مختلف موجود با یکدیگر ترکیب کرد تا هسته جدیدی به دست بیاید.
- نکته‌ی شایان توجه دیگر این که با توجه دو روشی که برای برچسب گذاری نقاط ذکر شد، این طور به نظر می‌رسد که در کاربرد واقعی، داده‌های برچسب‌دار ما در فاز آموزش، تعدادی پیکسل پراکنده نیستند، بلکه طبعاً به این صورت است که فرد متخصص، تعدادی ناحیه - که هر ناحیه شامل تعدادی پیکسل است - را مشخص کرده و برچسب مربوط به هر ناحیه را می‌گوید.
- چندکلاسه بودن مساله: بسیاری از روش‌های دسته بندی، برای مسائل دو کلاسه ارائه شده‌اند. بنابراین برای تعمیم به حالت چندکلاسه، مجبور به استفاده از روش‌هایی مثل دسته‌بندی یک با یک^{۱۶} یا دسته‌بندی یک با همه^{۱۷} می‌باشیم. البته در روش‌های نیمه‌نظارتی، چون از داده‌های بدون برچسب هم در

یادگیری استفاده می‌شود، استفاده از روش دسته‌بندی یک با یک چندان درست نیست، زیرا در این صورت، برای تمایز بین دو کلاس، از داده‌های بدون برچسب مربوط به سایر کلاس‌ها هم به عنوان داده‌های مربوط به این دو کلاس استفاده خواهیم کرد.

۳ استخراج ویژگی

همان‌طور که در بخش ۱ گفته شد، زیاد بودن تعداد ابعاد، دقت دسته‌بندی را کاهش می‌دهد. همچنین این مساله، در حجم داده‌هایی که باید ذخیره و احتمالاً به مقصدی ارسال شوند هم ایجاد مشکل می‌کند. یعنی اگر روش خوبی برای کاهش بعد داده‌ها وجود داشته باشد، می‌توان داده‌ها را پس از عکس‌برداری ابتدا کاهش بعد داد، و سپس از ماهواره به زمین ارسال کرد.

بنابراین برای سنجش میزان کارایی یک روش کاهش بعد، دو معیار وجود دارد، یکی تعداد ابعاد لازم برای ذخیره‌ی مقدار مورد نظر از انرژی داده‌ها، و دیگری تفکیک‌پذیری داده‌ها پس از کاهش بعد. معیار اول در روش‌هایی مثل تحلیل مولفه‌ی اصلی^{۱۸} و آیزومپ^{۱۹}، واریانس باقیمانده^{۲۰} است که نسبت واریانس نقاط در فضای جدید به واریانس آن‌ها در فضای اولیه (نسبت مجموع مقادیر ویژه‌ی مربوط به ابعاد مورد استفاده، به مجموع تمام مقادیر ویژه) می‌باشد.

اما تفکیک‌پذیری داده‌ها معمولاً با سنجش دقت یک دسته‌بند ساده، مثلاً دسته‌بند مبتنی بر k -نزدیک‌ترین همسایه، روی داده‌های کاهش بعد داده شده به روش‌های مختلف مقایسه می‌شود.

در [۲] بر اساس این دو معیار نشان داده شده است که آیزومپ روش خوبی برای کاهش بعد داده‌های منیفلدی می‌باشد، و در مقالات دیگر از این مطلب به عنوان شاهی برای این که داده‌های سنجش از دور، بر روی منیفلدی با تعداد ابعاد کم قرار دارند استفاده شده است. نویسندگان این مقاله، در مقالات بعدی در جهت بهبود زمان اجرا و حافظه‌ی مورد نیاز روش آیزومپ تلاش کرده‌اند، و نهایتاً در [۳]، روشی تخمینی برای آیزومپ ارائه کرده‌اند که زمان اجرای آن از مرتبه‌ی $O(n \lg^2 n)$ و حافظه‌ی مورد نیاز آن از مرتبه‌ی $O(n \lg n)$ می‌باشد (n تعداد داده‌ها، یعنی همان تعداد پیکسل‌ها است).

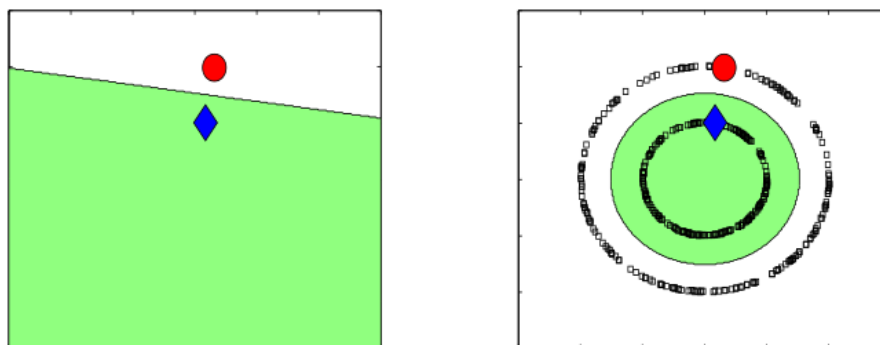
از دیگر روش‌های معروف کاهش بعد مبتنی بر منیفلد، می‌توان تعبیه‌ی موضعی خطی^{۲۱}، تنظیم موضعی صفحه‌ی مماس^{۲۲} و نگاشت ویژه‌ی لاپلاسی^{۲۳} را نام برد. این روش‌ها خواص موضعی منیفلد را حفظ می‌کنند، یعنی سعی می‌کنند موقعیت نقاط نزدیک به یکدیگر در فضای اولیه، در فضای دوم نیز حفظ شود. این روش‌ها نیز در مقالات مختلفی مثل [۴، ۵، ۶، ۷] بررسی شده‌اند.

ما در این پژوهش بر روش‌های کاهش بعد تمرکز نخواهیم کرد، زیرا هدف نهایی ما به دست آوردن دقت کافی در دسته‌بندی داده‌هاست. در ضمن با استفاده از روش‌های مبتنی بر هسته، با اتخاذ هسته مناسب می‌توان از افزایش خطا به دلیل تعداد زیاد ابعاد جلوگیری کرد. در مورد زمان اجرا و حافظه‌ی مورد نیاز هم، در اکثر روش‌های دسته‌بندی، تعداد زیاد داده‌ها عامل اصلی زیاد بودن زمان و حافظه‌ی مورد نیاز می‌باشد، نه تعداد ابعاد زیاد داده‌ها. نهایتاً در صورتی که در روش خود نیاز به استفاده از کاهش بعد داشتیم، می‌توانیم از روش‌هایی که تا کنون مطرح شده‌اند استفاده کنیم.

۴ یادگیری نیمه‌نظارتی

در یادگیری نظارتی^{۲۴} آموزش به کمک داده‌هایی انجام می‌شود که برچسب تمام آن‌ها مشخص است. به عبارت دیگر در فاز آموزش تلاش می‌شود تا به کمک داده‌های برچسب‌دار و توزیع آن‌ها، ساختاری پیدا شود که بر اساس آن بتوانیم در فاز تست برچسب داده‌های تست را تعیین کنیم. اما در یادگیری نیمه‌نظارتی^{۲۵} در فاز آموزش، علاوه بر داده‌های برچسب‌دار، تعدادی داده‌ی بدون برچسب هم در اختیار داریم که در تشخیص نحوه‌ی پراکندگی نقاط می‌توانیم از آن‌ها کمک بگیریم. به عنوان مثالی ساده فرض کنید برای دسته‌بندی داده‌های نشان داده شده در شکل ۱.۴ به دو دسته‌ی آبی و قرمز، بخواهیم مرزی پیدا کنیم که بر اساس آن نقاط صفحه را به دو دسته نسبت دهیم. اگر بخواهیم فقط از اطلاعات نقاط برچسب‌دار استفاده کنیم، احتمالاً مرز را شبیه آن‌چه در شکل سمت چپ نشان داده شده است تعیین می‌کنیم. اما اگر داده‌های بدون برچسب را هم در اختیار داشته باشیم، مرزی که انتخاب می‌کنیم، چیزی شبیه به شکل سمت راست خواهد بود.

استفاده از یادگیری نیمه‌نظارتی، مخصوصاً در مساله‌هایی مثل مساله‌ی ما که در آن‌ها برچسب‌گذاری دستی داده‌ها کار آسانی نیست و بنابراین تعداد داده‌های برچسب‌دار موجود کم است، همچنین تعداد زیادی داده‌ی بدون برچسب در اختیار داریم، بسیار مفید می‌باشد.



شکل ۱.۴: استفاده از داده‌های بدون برچسب در یادگیری [۸]

در یادگیری نیمه‌نظارتی، ممکن است هدف نهایی، برچسب‌گذاری همان داده‌های بدون برچسبی باشد که در فاز آموزش در اختیار داشتیم. این نوع یادگیری را یادگیری ورارسانی^{۲۶} می‌گویند. در حقیقت این نوع یادگیری خیلی قابل تفکیک به دو فاز آموزش و تست نمی‌باشد.

اما نوع دیگر یادگیری نیمه‌نظارتی که آن را یادگیری استنتاجی^{۲۷} می‌گویند به این صورت است که در فاز تست، باید برچسب داده‌هایی که در فاز آموزش اصلاً در اختیار ما نبودند را مشخص کنیم.

در ساده‌ترین صورت مساله‌ی ما، هدف این است که برچسب همان داده‌هایی را که در فاز یادگیری به صورت بدون برچسب در اختیار داشتیم تعیین کنیم، یعنی یادگیری ورارسانی. در واقع در این نوع یادگیری تمام کار ما با یک عکس ابرطیفی است که برچسب بعضی از پیکسل‌های آن را می‌دانیم و می‌خواهیم برچسب سایر پیکسل‌ها را مشخص کنیم.

همچنین هدف ما در یادگیری می‌تواند به این صورت باشد که با استفاده از یک یا چند عکس، یادگیری را انجام دهیم و از آن به بعد، عکس‌هایی که از مناطق مختلف برداشته می‌شود را برچسب‌گذاری کنیم. در این حالت، با یک مساله‌ی یادگیری استنتاجی مواجه خواهیم بود.

مساله‌ی مهمی که در سنجش از دور مطرح است این است که الگوهای موجود برای یک برچسب (مثلاً یک نوع پوشش گیاهی یا یک نوع بافت سطح زمین) در عکس‌هایی که از مناطق مختلف گرفته می‌شوند، یا حتی در قسمت‌های مختلف یک عکس، متفاوت می‌باشند. برای مثال در مساله‌ی تشخیص مناطق شهری، نوع مصالح و روش ساختمان‌سازی و جاده‌سازی شهرهای مختلف می‌تواند باعث تفاوت این الگوها باشد. بنابراین لازم است روش پیشنهادی به صورتی باشد که بتواند از دانش فراگرفته شده در یادگیری یک عکس یا یک قسمت از عکس، برای یادگیری یا تست در عکس‌های دیگر یا قسمت‌های دیگر همان عکس استفاده کند. همچنین به دلیل زیاد بودن تعداد داده‌ها، ممکن است بخواهیم به جای در نظر گرفتن مجموع عکس‌های قدیمی و عکس‌هایی که تازه به دستمان رسیده، به عنوان داده‌های آموزش، از دانش به‌دست آمده در تحلیل عکس‌های قبلی، به نحوی در تحلیل عکس‌های جدید استفاده کنیم. به همین دلیل، برخی تلاش‌ها در جهت ارائه‌ی چارچوبی برای یادگیری است که قابلیت انتقال دانش^{۲۸} داشته باشد. انتقال دانش هم در یادگیری استنتاجی و هم در یادگیری ورارسانی کاربرد دارد.

۵ روش‌های یادگیری نیمه‌نظارتی به کار رفته در سنجش از دور

به طور کلی روش‌های یادگیری نیمه‌نظارتی مورد استفاده در سنجش از دور را می‌توان به سه دسته‌ی روش‌های مولد^{۲۹}، روش‌های مبتنی بر جداسازی در ناحیه‌ی کم تراکم و روش‌های مبتنی بر گراف تقسیم کرد. چون منیفلدها معمولاً با گراف تخمین زده می‌شوند، روش‌های منیفلدی همان روش‌های مبتنی بر گراف می‌باشند. البته روش‌های مبتنی بر گراف را هم می‌توان به صورت حالتی خاص و یا تعمیمی از دو روش دیگر دانست، اما برای بررسی و تمایز بیشتر، در قسمتی جداگانه به بررسی آن‌ها می‌پردازیم.

۱.۵ روش‌های مولد

روش‌های مولد به این صورت هستند که در قسمتی از یادگیری خود، از تخمین $P(x|y)$ ، یعنی توزیع داده برحسب برچسب آن، استفاده می‌کنند، یعنی این که این روش‌ها برای هر کلاس، سعی می‌کنند توزیع داده‌های مربوط به آن کلاس را تخمین بزنند. البته طبعاً این تخمین براساس فرض اولیه‌ای بر روی توزیع داده‌های هر کلاس می‌باشد، مثلاً این که نوع این توزیع گاوسی می‌باشد.

از روش‌های مولد نیمه‌نظارتی معروف، روش EM^{۳۰} می‌باشد. این روش در [۹] بر روی داده‌های سنجش از دور اعمال شده است. البته کم بودن تعداد داده‌های برچسب‌دار، می‌تواند در تخمین توزیع اولیه‌ای که روش EM کار خود را از آن آغاز می‌کند، خطایی ایجاد کند که نتیجه‌ی آن رسیدن به جواب نهایی اشتباه باشد. به طور کلی، روش‌های مولد در مسائلی که تعداد داده‌های برچسب‌دار کم است، با چنین مشکلی مواجه می‌باشند. بنابراین به نظر می‌رسد با استفاده از این روش‌ها، تخمین خوبی برای توزیع داده‌های هر کلاس به دست نیاید.

ضمناً در نقد این روش، این سوال که چرا نمایش طیفی داده‌های مربوط به هر کلاس از توزیعی خاص، مثلاً گاوسی، پیروی می‌کند هم قابل طرح می‌باشد.

۲.۵ روش‌های مبتنی بر جداسازی در ناحیه‌ی کم تراکم

همان‌طور که در قسمت ۱.۵ گفته شد، وقتی داده‌های برچسب‌دار کمی در اختیار داریم، استفاده از روش‌های مولد نمی‌تواند دسته‌بندی را با دقت خوبی انجام دهد. در این گونه موارد معمولاً روش‌های مبتنی بر جداسازی در ناحیه‌ی کم تراکم، از جمله ماشین بردار پشتیبان^{۳۱} می‌تواند راهگشا باشد. در عمل هم دیده می‌شود که نتایج حاصل از این روش، حتی با هسته خطی، خیلی خوب است. برای استفاده از داده‌های بدون برچسب و یادگیری نیمه‌نظارتی، می‌توان از ماشین بردار پشتیبان ورارسانی^{۳۲} استفاده کرد، که از این ایده در مقالاتی مثلاً [۱۰] استفاده شده است.

از مزایای روش ماشین بردار پشتیبان این است که به دلیل مبتنی بودن بر هسته، با زیاد شدن تعداد ابعاد دچار مشکل نمی‌شوند، فقط یادگیری پارامترهای هسته بهینه در آن‌ها می‌تواند کمی زمان‌بر باشد. همچنین می‌توانیم ویژگی‌های خاص داده‌های مساله را به صورت هسته خاصی که با توجه به شرایط مساله تعریف می‌کنیم، در یادگیری دخیل کنیم. مثلاً می‌توان ویژگی‌های مکانی را با دخیل کردن هسته وابسته به مکان پیکسل‌ها در مساله دخیل کرد.

همچنین می‌توان از داده‌های بدون برچسب، بوسیله‌ی تاثیر آن‌ها در هسته مورد استفاده بهره برد. برای مثال در [۱۱] از داده‌های بدون برچسب به این صورت استفاده شده است که با استفاده از آن‌ها، عمل خوشه‌بندی به روش k-میانگین^{۳۳} را چند بار، با شروع از حالات اولیه‌ی متفاوت انجام می‌دهد، و با توجه به این که دو نقطه در چند بار اجرای خوشه‌بندی در یک خوشه قرار گرفته‌اند، معیار شباهتی تعریف می‌کند و این هسته را با هسته گاوسی ترکیب کرده و در ماشین بردار پشتیبان استفاده می‌کند.

همچنین در مسائلی مثل تشخیص مناطق شهری که هدف تشخیص یک کلاس می‌باشد، روش تشخیص دامنه‌ی بردار پشتیبان^{۳۴} هم کاربرد دارد، مثلاً در [۱۲، ۱۳] از این ایده استفاده شده است. در این مقالات همچنین روش‌هایی برای تعمیم روش‌های دسته‌بندی یک کلاس به دسته‌بندی چندکلاس در سنجش از دور ارائه شده است.

در روش‌های مبتنی بر ماشین بردار پشتیبان، می‌توان از هسته‌های مبتنی بر منیفلد استفاده کرد. در مورد این ایده در بخش ۶ توضیح خواهیم داد.

۳.۵ روش‌های مبتنی بر گراف

در روش‌های منیفلدی، هر داده - اعم از داده‌های برچسب دار و بدون برچسب - به صورت یک راس در فضای ویژگی‌ها در نظر گرفته می‌شود. منیفلد رویه‌ای است که به صورت موضعی خواص فضای خطی را دارد، بنابراین فاصله‌ی اقلیدسی دو نقطه‌ی نزدیک به هم تخمین خوبی برای فاصله‌ی آن‌ها در روی منیفلد است. به همین دلیل است که برای تخمین فاصله‌ی ژئودزیک دو نقطه (یعنی فاصله‌ی آن‌ها روی منیفلد)، از فاصله‌ی آن‌ها روی گراف k -نزدیک‌ترین همسایگی استفاده می‌کنند.

روش‌های یادگیری مبتنی بر منیفلد، بر اساس دو فرض استوار می‌باشند، نخست این که داده‌ها روی منیفلدی هستند که با گراف توضیح داده شده تخمین زده می‌شود، و دوم این که تابع برچسب گذاری بر روی این منیفلد هموار است و به طور میانگین، تغییرات محلی کمی دارد.

استفاده از منظم‌سازی تیخونوف

همان‌طور که گفته شد، در روش‌های دسته‌بندی (و رگرسیون) مبتنی بر منیفلد دو فرض وجود دارد، یکی این که داده‌ها بر روی یک منیفلد با بعدی کمتر از بعد فضای ویژگی قرار دارند، و دوم این که تغییرات تابع برچسب‌گذاری بر روی منیفلد هموار است، یعنی اختلاف برچسب نقاط در روی نقاط نزدیک به هم، به طور متوسط کم می‌باشد. با استفاده از این دو فرض، می‌توان یک مساله‌ی بهینه‌سازی پیشنهاد کرد که به نوعی معادل مساله‌ی یادگیری مبتنی بر منیفلد است. عبارتی که باید بر حسب تابع برچسب‌گذاری (f) بهینه شود دارای دو جمله می‌باشد، یکی خطای دسته‌بندی داده‌های برچسب‌دار را نشان می‌دهد، و دیگری میزان همواری تابع برچسب‌گذاری را بر روی منیفلد. برای سنجش میزان همواری تابع برچسب‌گذاری بر روی منیفلد، در میزان همواری آن در برچسب‌گذاری رئوس نزدیک به هم در گراف سنجیده می‌شود. اگر W را به عنوان ماتریس شباهت داده‌ها در نظر بگیریم (یعنی W_{ij} میزان شباهت دو راس i و j را نشان دهد) آنگاه $\sum_{i,j} W_{ij}(f_i - f_j)^2$ معیار خوبی برای سنجش همواری تابع برچسب‌گذاری می‌باشد (f_i برچسب نسبت داده شده به راس i در یک مساله‌ی دسته‌بندی دو کلاسه است). این همان معیاری است که در روش منظم‌سازی تیخونوف^{۳۵} برای تضمین همواری تابع وجود دارد. در حقیقت عبارت بالا بر این نکته تاکید می‌کند که هرچه دو نقطه به یکدیگر نزدیک‌تر هستند، باید برچسب‌های مشابه‌تری داشته باشند. W به صورت یک کرنل RBF تعریف می‌شود:

$$W_{ij} = \exp(-\|x_i - x_j\|^2 / 2\delta^2) \quad (۱)$$

البته برای تخمین بهتری از منیفلد، می‌توان برای هر راس، فقط k نزدیک‌ترین راس را به عنوان همسایه‌های آن در نظر گرفت.

مفهوم مهمی با نام لاپلاسین گراف به صورت

$$L = D - W \quad (۲)$$

تعریف می‌شود. با این تعریف می‌توان نشان داد که:

$$\sum_{i,j} W_{ij}(f_i - f_j)^2 = f^T L f \quad (۳)$$

بنابراین، عبارتی که باید بر حسب تابع برچسب‌گذاری بهینه شود، به این صورت می‌باشد:

$$\gamma f^T L f + \frac{1}{l} \sum_{i=1}^l (f_i - y_i)^2 \quad (۴)$$

که y_i برچسب واقعی داده‌ی i است که در فاز آموزش، به صورت برچسب‌دار در اختیار ما می‌باشد (یعنی طبیعتاً مقدار خطا، فقط روی داده‌های برچسب‌دار فاز آموزش محاسبه می‌شود).

روش سازگاری محلی و عمومی

یکی از روش‌های دسته‌بندی مبتنی بر منیفلد، روش سازگاری محلی و عمومی^{۳۶} می‌باشد. در [۱۴] ایده‌ی استفاده از این روش برای تصاویر ابرطیفی سنجش از دور مطرح شده است، البته در این مقاله، از ترکیب هسته مکانی با هسته گاوسی مورد استفاده در روش سازگاری محلی و عمومی هم استفاده می‌شود.

ایده‌ی اصلی روش سازگاری محلی و عمومی این است که رئوس در هر مرحله برچسب خود را به همسایه‌های خود انتشار می‌دهند، تا به حالتی پایدار برسند. این انتشار برچسب و رسیدن به حالت پایدار، در حقیقت ضامن همان همواری تابع برچسب‌گذاری بر روی منیفلد می‌باشد. مقدار انتشار برچسب راس i به راس j ، متناسب با وزن یال بین دو راس، یعنی شباهت بین دو راس (W_{ij}) می‌باشد. معمولاً W_{ij} به صورت یک کرنل RBF تعریف می‌شود:

$$W_{ij} = \exp(-\|x_i - x_j\|^2 / 2\delta^2) \quad (۵)$$

البته با این استثنا که $W_{ii} = 0$.

در ادامه توضیح روش سازگاری محلی و عمومی آمده است. فرض کنید پیکسل‌های N بعدی به صورت $X = \{x_1, \dots, x_l, x_{l+1}, \dots, x_n\} \in R^N$ باشند که برچسب l تای اول آن‌ها مشخص و برچسب بقیه نامشخص است. مجموعه‌ی برچسب‌های مختلف را هم $L = \{1, \dots, c\}$ در نظر بگیرید. فرض کنید \mathcal{F} مجموعه‌ی تمام ماتریس‌های $n \times c$ با درایه‌های غیر منفی باشد. از روی ماتریسی مثل $F = [F_1^T, \dots, F_n^T]^T \in \mathcal{F}$ می‌توان هر راس x_i را با برچسب $y_i = \arg \max_{j \leq c} F_{ij}$ برچسب‌گذاری کرد. بنابراین می‌توان گفت که F تابعی از X به R^c است که به هر راس، برداری از c عدد غیر منفی نسبت می‌دهد. این اعداد قرار است کارکردی شبیه به احتمال اتخاذ هر برچسب توسط هر راس را داشته باشند، یعنی با کمی اغماض می‌توان گفت که در پایان الگوریتم، بزرگتر بودن درایه‌ی $F_{i,a}$ از $F_{i,b}$ ، نشان‌دهنده‌ی بیشتر بودن احتمال اتخاذ برچسب a توسط راس i ، در مقایسه با برچسب b است.

همچنین ماتریس Y با ابعاد $n \times c$ را به این صورت تعریف می‌کنیم که $Y_{i,j} = 1$ اگر برچسب x_i برابر با j باشد، وگرنه $Y_{i,j} = 0$. بدین ترتیب Y در حقیقت متناسب با برچسب‌گذاری ابتدایی داده‌ها می‌باشد.

بدین ترتیب الگوریتم دسته‌بندی به صورت زیر بیان می‌شود:

۱. ماتریس وزن یال‌ها (W) را محاسبه می‌کنیم.

۲. ماتریس $S = D^{-1/2} W D^{-1/2}$ را تشکیل می‌دهیم، که در آن W ماتریس شامل درایه‌های W_{ij} است که در بالا تعریف شد، و D ماتریسی قطری است که درایه‌ی (i, i) آن برابر با جمع درایه‌های سطر i ام W می‌باشد.

۳. عمل زیر را تا زمان همگرایی تکرار می‌کنیم:

$$F(t+1) = \alpha S F(t) + (1 - \alpha) Y$$

۴. در نهایت برچسب داده‌ها به این صورت تعیین می‌شود:

$$y_i = \operatorname{argmax}_{j \leq c} F_{ij}^*$$

دلیل این که در قسمت ۱ از الگوریتم بالا، قرار می‌دهیم $W_{ii} = 0$ این است که هیچ راسی برچسبش را دوباره به خودش منتقل نکند. مرحله‌ی ۲ برای نرمال‌سازی وزن‌هاست، این کار تضمین‌کننده‌ی همگرایی الگوریتم به حالتی پایدار است. در قسمت ۳ از الگوریتم، α پارامتری در بازه‌ی $(0, 1)$ می‌باشد که مشخص می‌کند هر راس، برای تعیین برچسبش در لحظه‌ی بعد، چه‌قدر از برچسب سایر رئوس تاثیر می‌پذیرد و چه‌قدر مایل به حفظ برچسب اولیه‌ی خودش می‌باشد.

تعبیر دیگری برای این روش این است که در ابتدا برای داده‌های برچسب‌دار، برچسب آن راس به او داده شده، و برای سایر داده‌ها برچسب صفر اتخاذ شده و سپس در هر مرحله، هر راسی برچسبش را متناسب با وزن یال‌های متصل به آن، به همسایه‌هایش منتقل می‌کند.

در [۱۵] نشان شده است که مقدار نهایی F از رابطه‌ی زیر به دست می‌آید:

$$F^* = \lim_{t \rightarrow \infty} F(t) = (I - \alpha S)^{-1} Y \quad (۶)$$

بنابراین برای محاسبه‌ی مقدار نهایی F ، فقط نیاز به محاسبه‌ی $(I - \alpha S)^{-1}$ داریم. یعنی محاسبه‌ی وارون یک ماتریس $n \times n$. برای این کار در این مقاله پیشنهاد شده است که از روش Nyström استفاده شود. این روش، در زمان کمتری، تخمینی از وارون ماتریس ارائه می‌کند.

بیان روش سازگاری محلی و عمومی به صورت مساله‌ی بهینه سازی

در [۱۵] نشان داده شده است که روش سازگاری محلی و عمومی معادل با حل مساله‌ی بهینه سازی $F^* = \arg \min_{F \in \mathcal{F}} \mathcal{Q}(F)$ می‌باشد که در آن:

$$\mathcal{Q}(F) = \frac{1}{2} \left(\sum_{i,j=1}^n W_{ij} \left\| \frac{1}{\sqrt{D_{ii}}} F_i - \frac{1}{\sqrt{D_{jj}}} F_j \right\|^2 + \mu \sum_{i=1}^n \|F_i - Y_i\|^2 \right) \quad (۷)$$

که $\mu > 0$ پارامتر منظم سازی^{۳۷} می باشد.

جمله‌ی اول در عبارت سمت راست تساوی بالا، وظیفه‌ی تامین همواری تابع برچسب گذاری بر روی منیفلد را دارد و جمله‌ی دوم، مقدار خطای تابع برچسب گذاری را نشان می دهد.

در حقیقت، این روش، در حالت چند کلاسه بیان شده، و روش حل مشکل وجود چند کلاس در آن، به روش دسته بندی یک با همه بوده است. یعنی حاصل حل تعدادی مساله‌ی دو کلاسه می باشد، که هر کدام به صورت مساله‌ی بهینه سازی عبارتی به این شکل می باشد:

$$\frac{1}{2} \sum_{i,j} W_{ij} \left(\frac{f_i^{(k)}}{\sqrt{D_{ii}}} - \frac{f_j^{(k)}}{\sqrt{D_{jj}}} \right)^2 + \frac{1}{2} \mu \sum_{i=1}^n \|f_i^{(k)} - Y_i(k)\|^2 \quad (8)$$

که در آن $f_i^{(k)}$ برچسبی است که به راس i می دهیم، وقتی که در حال دسته بندی کلاس k در مقابل تمام کلاس های دیگر هستیم، یعنی زمانی که با یک مساله‌ی بهینه سازی دو کلاسه، بر حسب تابع برچسب گذاری $f^{(k)}$ مواجه هستیم.

مفهومی با نام لاپلاسین نرمال شده به این صورت تعریف می شود:

$$L_s = D^{-1/2} L D^{-1/2} \quad (9)$$

همان طور که قبلا گفته شد، جمله‌ی $f^T L f$ برای سنجش میزان همواری تابع بر روی منیفلد به کار می رود. با استفاده از لاپلاسین نرمال شده به جای لاپلاسین عادی در این عبارت خواهیم داشت:

$$f^T L_s f = \frac{1}{2} \sum_{i,j} W_{ij} \left(\frac{f_i}{\sqrt{D_{ii}}} - \frac{f_j}{\sqrt{D_{jj}}} \right)^2 \quad (10)$$

این جمله همان جمله‌ای است که در معادله‌ی ۸ برای تضمین همواری تابع به کار رفته است. بنابراین تفاوت دو روش سازگاری محلی و عمومی و منظم سازی تیخونوف در قسمت عبارت مربوط به همواری تابع، در نوع لاپلاسین به کار رفته است. وقتی از لاپلاسین نرمال شده استفاده می کنیم، در حقیقت میزان همواری تابع در هر نقطه را با توجه به میزان تراکم در آن جا می سنجیم.

تفاوت دیگر روش سازگاری محلی و عمومی با منظم سازی تیخونوف در محاسبه‌ی خطای تابع برچسب گذاری است. جمله‌ی دوم در عبارت سمت راست معادله‌ی ۸ برای محاسبه‌ی خطای تابع برچسب گذاری F ، برچسب داده شده به وسیله‌ی F را با برچسب اولیه‌ی رئوس مقایسه می کند، اما این کار را برای تمام رئوس، اعم از برچسب دار و بدون برچسب، انجام می دهد. این در صورتی است که برچسب اولیه‌ی نقاط بدون برچسب، معنای چندانی ندارد، و چون مقدار اصلی برچسب آن ها را نمی دانیم، نباید در محاسبه‌ی خطای تابع برچسب گذاری آن ها را در نظر گرفت. در روش منظم سازی تیخونوف، مقدار این خطا فقط بر روی رئوس برچسب دار محاسبه می شود.

روش ماشین بردار پشتیبان لاپلاسینی

روش های مبتنی بر منیفلد که تا کنون مطرح کردیم، به صورت یک مساله‌ی بهینه سازی قابل طرح بودند. عباراتی که باید بهینه سازی می شدند از دو جمله تشکیل شده بودند: تابع ضرر^{۳۸} و جمله‌ی ضامن همواری تابع بر روی منیفلد. منظور از تابع ضرر همان تابعی است که خطای برچسب دهی به نقاط برچسب دار آموزش را حساب می کند.

روش های غیر منیفلدی، جمله‌ی ضامن همواری تابع بر روی منیفلد را ندارند، اما جمله‌ی دیگری دارند که میزان اطمینان دسته بندی را زیاد می کند، و آن اندازه‌ی تابع در فضای هیلبرت بازتولید هسته^{۳۹} می باشد. مثلا در ماشین بردار پشتیبان، جمله‌ی $\frac{1}{2} \|w\|^2$ همان اندازه‌ی تابع در فضای هیلبرت بازتولید هسته می باشد.

روش ماشین بردار پشتیبان لاپلاسی، با اضافه کردن این جمله به عبارت بهینه‌سازی روش‌های منیفلدی، و همچنین با در نظر گرفتن تابع لولا^{۴۰} به عنوان تابع ضرر به وجود آمده است. تابع ضرر لولا، همان تابع ضرر مورد استفاده در ماشین بردار پشتیبان می‌باشد. بنابراین عبارتی که باید بهینه‌سازی شود به صورت

$$\frac{1}{l} \sum_{i=1}^l \max(\cdot, 1 - y_i \cdot f_i) + \gamma_L \|f\|_K^2 + \gamma_M f^t L f \quad (11)$$

می‌باشد.

در عمل هم دیده شده که این روش از دقت خوبی در دسته‌بندی تصاویر سنجش از دور برخوردار است، استفاده از این روش در سنجش از دور، در مقالاتی مثل [۱۶، ۱۷، ۱۸، ۱۹] مطرح شده است.

از مزایای این روش نسبت به دو روش قبل این است که تابع برچسب‌دهی را در تمام نقاط فضا تخمین می‌زند، نه فقط نقاط آموزش و تست. به عبارتی، یادگیری با این روش، یادگیری استنتاجی است، در صورتی که در روش‌های قبلی، یادگیری به صورت ورارسانی می‌باشد. در جدول ۱.۵ خلاصه‌ای از روش‌های نیمه‌نظارتی به کار رفته در سنجش از دور که در مورد آن‌ها توضیح دادیم، ارائه شده است.

۶ مقایسه‌ی عملی روش‌های دسته‌بندی

برای مقایسه‌ی روش‌های مختلف دسته‌بندی مبتنی بر منیفلد که در بخش ۳.۵ به آن‌ها اشاره شد، آن‌ها را روی مجموعه داده‌ی یکسانی اعمال کرده و نتایج را با یکدیگر مقایسه می‌کنیم. مجموعه داده‌ی مورد استفاده، تصویری است که در پروژه‌ی آپیریس^{۴۱}، زیر نظر سازمان ملی هوافضای آمریکا (ناسا^{۴۲})، در سال ۱۹۹۲ از سایت تست ایندین پاین^{۴۳} گرفته شده است و در [۲۰] قابل دسترسی می‌باشد. داده‌های این تصویر، مربوط به ۱۶ نوع پوشش گیاهی می‌باشند. تعداد داده‌های مربوط به هر کلاس، در جدول ۲.۶ نمایش داده شده است.

وقتی تعداد داده‌های کلاس‌های مختلف خیلی متفاوت باشند، نمی‌توان دقت دسته‌بندی را به درستی سنجید، به همین دلیل فقط از کلاس‌های ۳ تا ۶، ۸، ۱۰، ۱۲، ۱۳، ۱۵ و ۱۶ استفاده شده است. تعداد داده‌های این دسته‌ها در مجموع ۵۰۰۷ داده است. همان‌طور که در [۲۱] گفته شده، باندهای [۱۰۴:۱۰۸] و [۱۵۰:۱۶۳] به دلیل تاثیرات ناشی از جذب آب اتمسفر، نویزی می‌باشند. بنابراین این باندها را حذف کردیم.

برای آموزش، از ۱۰ درصد داده‌ها به عنوان داده‌های برچسب‌دار آموزش استفاده شد، به این صورت که در هر بار اجرا، از هر کلاس، ۱۰ درصد پیکسل‌ها به صورت تصادفی به عنوان داده‌ی برچسب‌دار، و سایر پیکسل‌ها به عنوان داده‌ی بدون برچسب انتخاب می‌شدند. هر روش، پنج مرتبه اجرا شده و میانگین و واریانس دقت دسته‌بندی داده‌های تست در این پنج اجرا محاسبه شده است. با توجه به تعداد زیاد ابعاد داده‌ها، در ابتدا به روش تحلیل مولفه‌ی اصلی، بعد داده‌ها به ۲۰ کاهش یافته است. به این ترتیب انتظار می‌رود که اثرات ناشی از افزایش بعد تا حدی از بین برود. نتیجه‌ی اجرای الگوریتم‌های مختلف، در جدول ۳.۶ نشان داده شده است.

می‌دانیم که دو روش یادگیری فرایند گاوسی^{۴۴} و ماشین بردار پشتیبان هر دو قابل بیان به صورت مساله‌ی بهینه‌سازی عبارتی شامل دو جمله هستند، که یک جمله اندازه‌ی تابع برچسب‌دهی در فضای هیلبرت بازتولید هسته، و دیگری مقدار خطای برچسب‌دهی به نقاط برچسب‌دار فاز آموزش می‌باشد. یعنی به صورت عبارت زیر:

$$OPT_f \left\{ \frac{1}{l} \sum_{i=1}^l v(f_i, y_i) + \gamma \|f\|_K^2 \right\} \quad (12)$$

که در آن $v(f_i, y_i)$ میزان خطا (تابع ضرر) در دسته‌بندی داده‌ی x_i می‌باشد، و $\|f\|_K^2$ اندازه‌ی تابع f در فضای هیلبرت بازتولید هسته مربوط به هسته مورد استفاده (K) می‌باشد. تفاوت دو روش در تابع ضرر مورد استفاده است. روش یادگیری فرایند گاوسی از تابع ضرر میانگین مربعات استفاده می‌کند ($v(f_i, y_i) = (f_i - y_i)^2$)، در صورتی که ماشین بردار پشتیبان از تابع ضرر لولا^{۴۵} استفاده می‌کند ($v(f_i, y_i) = \max(\cdot, 1 - y_i \cdot f_i)$).

وقتی در هر دو روش از یک هسته استفاده می‌کنیم، جمله‌ی مربوط به اندازه‌ی تابع در فضای هیلبرت بازتولید هسته در هر دو یکسان می‌باشد، اما تابع ضرر مورد استفاده‌ی آن‌ها متفاوت است. بنابراین از مقایسه‌ی نتایج به دست آمده از دو روش ماشین بردار پشتیبان با هسته خطی و یادگیری فرایند گاوسی با هسته

جدول ۱.۵: روش‌های دسته‌بندی نیمه‌نظارتی به کار رفته در سنجش از دور

مقالات	مزایا و معایب	توضیحات	روش‌های مولد
از این روش در مقالاتی مثل [۹] استفاده شده‌است.	<p>+ به دلیل ساختار تکرارشونده و خودآموز الگوریتم، می‌تواند از توزیع داده‌های بدون برچسب در تخمین پارامترهای مربوط به توزیع داده‌های هر کلاس استفاده کند.</p> <p>- مبتنی بر این فرض است که داده‌های هر کلاس دارای توزیع خاصی، مثلاً توزیع گاوسی، هستند.</p> <p>- در صورت کم بودن تعداد داده‌های برچسب‌دار، در تخمین توزیع داده‌ها اشتباه می‌کند، و با شروع از تخمین اولیه اشتباه، در نقطه‌ی بهینه‌ی موضعی اشتباهی متوقف می‌شود.</p>	روش‌هایی مثل EM که بر اساس تخمین توزیع داده‌های مربوط به هر کلاس در فضای ویژگی کار می‌کنند.	روش‌های مولد
برای مثال در [۱۰] از ماشین بردار پشتیبان و در [۱۲، ۱۳] از تشخیص دامنه‌ی بردار پشتیبان استفاده شده است. همچنین در مقالات زیادی، از روش‌های یادگیری فعال مبتنی بر ماشین بردار پشتیبان استفاده شده است. در مقالاتی مثل [۱۱]، ایده‌ی ترکیب هسته‌های دیگر (مثلاً هسته مبتنی بر خوشه‌بندی در این مقاله) با هسته طیفی و استفاده از آن در ماشین بردار پشتیبان مطرح شده است.	<p>+ به دلیل مبتنی بر هسته بودن این روش، می‌توان ویژگی‌های مکانی نقاط، یا ایده‌های منیفلدی را در قالب کرنل مربوطه، در یادگیری دخیل کرد.</p> <p>+ به دلیل مبتنی بر هسته بودن، از زیاد بودن تعداد ابعاد فضای ویژگی آسیب نمی‌بیند (البته به شرط انتخاب هسته مناسب). + به دلیل بی‌نیازی از تخمین توزیع داده‌های هر کلاس، با وجود تعداد کم داده‌های برچسب‌دار هم می‌تواند به خوبی کار کند.</p> <p>- در شکل ساده‌اش، فقط یک جداسازی خطی انجام می‌دهد. برای افزایش دقت، باید هسته مناسب برای داده‌های مساله را یافت.</p>	به طور خاص، روش ماشین بردار پشتیبان از دقت خوبی در دسته‌بندی تصاویر سنجش از دور برخوردار است. در کاربردهای یک‌کلاسه و یا حتی چندکلاسه، از روش‌های مبتنی بر تشخیص دامنه‌ی بردار پشتیبان هم استفاده شده است.	روش‌های مبتنی بر جداسازی در ناحیه‌ی کم‌تراکم
ابتدا استفاده از روش سازگاری محلی و عمومی در مقالاتی مثل [۱۴] و بعد از آن، استفاده از روش LapSVM در مقالاتی مثل [۱۶، ۱۷، ۱۸، ۱۹] مطرح شد. در این مقالات بحث‌های دیگری مثل ترکیب هسته مکانی با هسته طیفی و ارائه‌ی چارچوبی برای انتقال دانش نیز مطرح شده است.	<p>+ این روش‌ها با اضافه کردن حمله‌ی ضامن همواری تابع برچسب‌گذاری بر روی منیفلد، شکل منیفلدی پراکندگی نقاط را در نظر می‌گیرند.</p> <p>- محاسبه‌ی وارون ماتریس‌ها در مواقع مورد نیاز، عملی زمان‌گیر است.</p> <p>- روش سازگاری محلی و عمومی فقط در یادگیری واریاسی قابل استفاده است، البته روش ماشین بردار پشتیبان لاپلاسی این مشکل را ندارد.</p>	تا کنون دو روش سازگاری محلی و عمومی و ماشین بردار پشتیبان لاپلاسی در سنجش از دور مورد استفاده قرار گرفته است.	روش‌های مبتنی بر گراف (منیفلد)

جدول ۲.۶: تعداد داده‌های مربوط به هر دسته در مجموعه داده‌ی مورد بررسی

دسته:	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰	۱۱	۱۲	۱۳	۱۴	۱۵	۱۶
تعداد پیکسل‌های مربوطه:	۴۶	۱۴۲۸	۸۳۰	۲۳۷	۴۸۳	۷۳۰	۲۸	۴۷۸	۲۰	۹۷۲	۲۴۵۵	۵۹۳	۲۰۵	۱۲۶۵	۳۸۶	۹۳

جدول ۳.۶: مقایسه‌ی روش‌های مختلف دسته‌بندی

روش یادگیری	میانگین دقت	واریانس دقت
یادگیری فرایند گاوسی با هسته خطی	۷۲.۷۸٪	۰.۶۲
ماشین بردار پشتیبان با هسته خطی	۸۹.۹۹٪	۲.۰۵
یادگیری فرایند گاوسی با کرنل RBF	۹۲.۱۶٪	۱.۱۰
ماشین بردار پشتیبان با هسته RBF	۹۰.۸۱٪	۱.۷۶
سازگاری محلی و عمومی	۸۱.۵۸٪	۱.۰۹
ماشین بردار پشتیبان با کرنل انتشار لاپلاسی	۸۲.۲۸٪	۱.۱۷

خطی، به نظر می‌رسد که تابع ضرر به کار رفته در ماشین بردار پشتیبان، یعنی تابع لولا، تابع خوبی برای داده‌های ما می‌باشد. دلیل این امر می‌تواند این باشد که نقاط کلاس‌های مختلف دارای پراکندگی‌های متفاوتی در نقاط دور از مرز جداسازی هستند، و نباید نقاط دور از مرز جداسازی، در تعیین مکان مرز تاثیر داشته باشند، زیرا در این صورت، کلاسی که دارای نقاط بیشتری در نواحی دور از مرز جدا کننده می‌باشد، مرز را به سمت خود متمایل میکند و باعث افزایش تعداد داده‌هایی می‌شود که به اشتباه دسته‌بندی می‌شوند.

البته اگر به نتایج حاصل از همین دو روش با هسته RBF دقت کنید، می‌بینید که دقت هر دو روش تقریباً یکسان است، و حتی ماشین بردار پشتیبان کمی بدتر نتیجه داده است. این مشاهده البته نتیجه‌گیری ما در مورد تابع ضرر مناسب مساله را با شبهاتی مواجه می‌کند. به نظر می‌رسد برای توجیه این مطلب، باید نشان دهیم که هسته RBF طوری نقاط را به فضای بازنشانی^{۴۶} می‌برد، که در آن جداسازی بر اساس بیشترین حاشیه‌ی امنیت و جداسازی بر اساس کمترین مجموع مربعات فاصله از مرز، تقریباً یکسان می‌باشند.

از مقایسه‌ی نتایج روش سازگاری محلی و عمومی و روش یادگیری فرایند گاوسی با هسته خطی هم می‌توان نکته‌ی جالبی دریافت. این نکته از آن‌جا ناشی می‌شود که می‌توان نشان داد روش سازگاری محلی و عمومی تقریباً معادل با همان روش یادگیری فرایند گاوسی است، با هسته خاصی که تقریباً برابر با L^+ می‌باشد. L^+ شبه معکوس ماتریس لاپلاسی گراف می‌باشد (می‌توان نشان داد که لاپلاسی، یک ماتریس منفرد می‌باشد و معکوس‌پذیر نیست). چون L^+ معکوس‌پذیر نیست، نمی‌تواند یک ماتریس هسته باشد. در واقع اگر یک فرایند گاوسی با هسته $K = (L + \epsilon^2 I)^{-1}$ (این هسته را هسته انتشار لاپلاسی^{۴۷} می‌گویند) داشته باشیم، در صورتی که ϵ به صفر میل کند، تابع برچسب‌گذاری به تابع حاصل از روش سازگاری محلی و عمومی میل می‌کند. هسته انتشار لاپلاسی در حقیقت همان شباهت نقاط بر روی فضای منیفلدی است.

بنابراین از مقایسه‌ی روش سازگاری محلی و عمومی و روش یادگیری فرایند گاوسی با هسته خطی به نظر می‌رسد که هسته انتشار لاپلاسی کرنل خوبی برای مساله‌ی ماست. این نتیجه، به نحوی شاهی است بر این که در مساله‌ی ما، فرض اول یادگیری مبتنی بر منیفلد برقرار است، یعنی داده‌ها واقعاً بر روی منیفلد قرار دارند.

با کنار هم گذاشتن دو حدسی که زدیم، می‌توانیم بگوییم روشی که از تابع ضرر لولا و هسته انتشار لاپلاسی استفاده می‌کند، باید از دقت بالایی برخوردار باشد. برای این کار می‌توان از ماشین بردار پشتیبان (یا حتی ماشین بردار پشتیبان لاپلاسی) با هسته انتشار لاپلاسی استفاده کرد. اما در عمل مشاهده شد که روش ماشین بردار پشتیبان با هسته انتشار لاپلاسی از دقت چندان بالایی برخوردار نیست. فعلاً توجیهی برای این مساله نداریم.

تکمیل بررسی روش‌های مختلف دسته‌بندی و انتخاب مناسب‌ترین عبارت بهینه‌سازی برای استفاده در مسالهی دسته‌بندی مورد مطالعه	تا ۲۰ بهمن ماه سال ۹۰
بررسی هسته‌های مکانی مختلف و انتخاب هسته مناسب	تا ۱۰ اسفند ماه سال ۹۰
مطالعه‌ی روش‌های حل مسائل دسته‌بندی چندکلاسه	تا آخر اسفندماه سال ۹۰
ارائه‌ی روش‌هایی برای حل مشکل چندکلاسه بودن مساله، انتقال دانش و کاهش زمان حل مساله، در صورت امکان	تا آخر اردیبهشت ماه سال ۹۱
تکمیل پایان‌نامه‌ی و دفاع از تز کارشناسی ارشد	تا آخر خرداد ماه سال ۹۱

۷ پیشنهاد برای ادامه‌ی کار

- مطابق آن چه تا کنون در این پژوهش انجام شده، می‌توان روش‌های مختلف دسته‌بندی را به صورت مسائل بهینه‌سازی عبارات مختلف تبیین کرد. در ادامه‌ی پژوهش، برای کشف این مطلب که کدام توابع به عنوان تابع ضرر و هسته در دسته‌بندی تصاویر سنجش از دور موثرتر می‌باشند، و علل آن، به طور عمیق‌تر تلاش خواهد شد.
- با وجود این که وابستگی مکانی نقاط ویژگی خاصی در مساله هست که در خیلی از مقالات به آن اشاره شده است، اما در عمل معمولاً از روش‌های خیلی ساده‌ای، مثل میانگین‌گیری هر پیکسل با اطرافیان‌ش، برای بهره‌بردن از این ویژگی استفاده می‌شود. ارائه‌ی روشی بهتر برای استفاده از این ویژگی می‌تواند هدف پژوهش‌های بعدی باشد. در چهارچوب بیان مسالهی یادگیری به صورت یک مسالهی بهینه‌سازی، این ویژگی معمولاً به صورت یک هسته مکانی که با هسته اصلی ترکیب می‌شود، یا جمله‌ای که به عبارت تحت بهینه‌سازی اضافه می‌شود، اعمال می‌شود.
- هر چند بعضی از کاربردهای سنجش از دور، مربوط به دسته‌بندی‌های دو کلاسه است، اما در مسالهی دسته‌بندی که مورد نظر ماست، در حالت کلی با چند کلاس مواجه هستیم. در مجموعه داده‌های موجود هم معمولاً بین ۴ تا ۱۶ کلاس موجود می‌باشد. زیاد شدن تعداد کلاس‌ها، دقت دسته‌بندی را کاهش می‌دهد. معمولاً در روش‌های یادگیری نیمه‌نظارتی، به سادگی و بدون بررسی دلایل ریاضی، از روش دسته‌بندی یک با همه برای تبدیل یادگیری دوکلاسه به چندکلاسه استفاده می‌شود. این در صورتی است که شاید بتوان از توزیع داده‌های مربوط به کلاسهای مختلف، اطلاعات پیشین^{۴۸} در مورد کلاس‌ها به دست آورد، و در دسته‌بندی چندکلاسه از آن‌ها استفاده نمود. برای مثال در [۱۹] روشی ارائه شده که در آن با توجه به داده‌های برچسب‌دار آموزش، کلاس‌ها را به صورت سلسله مراتبی تقسیم‌بندی می‌کند، به طوری که برای حل مساله، در هر مرحله فقط لازم باشد دسته‌بندی را برای تمایز بین دو دسته از کلاسها انجام دهد.
- با وجود این که مفهوم انتقال دانش در سنجش از دور مطرح شده و برخی به آن پرداخته‌اند، تلاش‌های انجام شده، معمولاً از پشتوانه‌ای تئوری در تحلیل روش خود برخوردار نیستند. یک ایده برای ادامه کار می‌تواند مدل کردن تغییرات الگوهای موجود در تصاویر به صورت ریاضی‌وار، و ارائه‌ی روش‌های مناسبی برای انتقال دانش با توجه به مدل ارائه شده برای تغییرات، باشد.

در جدول ۴.۷ زمان‌بندی کلی برای ادامه‌ی پژوهش قابل مشاهده می‌باشد.

References

- [1] G. Camps-Valls, "Machine learning in remote sensing data processing," In *Machine Learning for Signal Processing, 2009. MLSP 2009. IEEE International Workshop on*, pp.1–6, sept. 2009.
- [2] C. Bachmann, T. Ainsworth, and R. Fusina, "Exploiting manifold geometry in hyperspectral imagery," *Geoscience and Remote Sensing, IEEE Transactions on*, Vol.43, pp.441–454, march 2005.

- [3] C. Bachmann, T. Ainsworth, and R. Fusina, "Improved manifold coordinate representations of large-scale hyperspectral scenes," *Geoscience and Remote Sensing, IEEE Transactions on*, Vol.44, pp.2786–2803, oct. 2006.
- [4] T. Han and D. Goodenough, "Nonlinear feature extraction of hyperspectral data based on locally linear embedding (lle)," In *Geoscience and Remote Sensing Symposium, 2005. IGARSS '05. Proceedings. 2005 IEEE International*, Vol.2, pp.1237–1240, july 2005.
- [5] L. Ma, M. Crawford, and J. Tian, "Generalised supervised local tangent space alignment for hyperspectral image classification," *Electronics Letters*, Vol.46, pp.497–498, 1 2010.
- [6] L. Ma, M. Crawford, and J. Tian, "Anomaly detection for hyperspectral images using local tangent space alignment," In *Geoscience and Remote Sensing Symposium (IGARSS), 2010 IEEE International*, pp.824–827, july 2010.
- [7] L. Ma, M. Crawford, and J. Tian, "Local manifold learning-based k-nearest-neighbor for hyperspectral image classification," *Geoscience and Remote Sensing, IEEE Transactions on*, Vol.48, pp.4099–4109, nov. 2010.
- [8] M. Belkin, P. Niyogi, and V. Sindhwani, "Manifold regularization: A geometric framework for learning from labeled and unlabeled examples," *The Journal of Machine Learning Research*, Vol.7, pp.2399–2434, 2006.
- [9] Q. Jackson and D. Landgrebe, "An adaptive classifier design for high-dimensional data analysis with a limited training data set," *Geoscience and Remote Sensing, IEEE Transactions on*, Vol.39, pp.2664–2679, dec 2001.
- [10] L. Bruzzone, M. Chi, and M. Marconcini, "A novel transductive svm for semisupervised classification of remote-sensing images," *Geoscience and Remote Sensing, IEEE Transactions on*, Vol.44, pp.3363–3373, nov. 2006.
- [11] D. Tuia and G. Camps-Valls, "Semisupervised remote sensing image classification with cluster kernels," *Geoscience and Remote Sensing Letters, IEEE*, Vol.6, pp.224–228, april 2009.
- [12] J. Munoz-Marf, L. Bruzzone, and G. Camps-Valls, "A support vector domain description approach to supervised classification of remote sensing images," *Geoscience and Remote Sensing, IEEE Transactions on*, Vol.45, pp.2683–2692, aug. 2007.
- [13] J. Munoz-Mari, G. Camps-Valls, L. Gomez-Chova, and J. Calpe-Maravilla, "Combination of one-class remote sensing image classifiers," In *Geoscience and Remote Sensing Symposium, 2007. IGARSS 2007. IEEE International*, pp.1509–1512, july 2007.
- [14] G. Camps-Valls, T. Bandos Marasheva, and D. Zhou, "Semi-supervised graph-based hyperspectral image classification," *Geoscience and Remote Sensing, IEEE Transactions on*, Vol.45, pp.3044–3054, oct. 2007.
- [15] D. Zhou, O. Bousquet, T. Lal, J. Weston, and B. Schölkopf, "Learning with local and global consistency," In *Advances in Neural Information Processing Systems 16: Proceedings of the 2003 Conference*, pp.595–602, 2004.
- [16] L. Gomez-Chova, G. Camps-Valls, J. Munoz-Mari, and J. Calpe, "Semi-supervised cloud screening with laplacian svm," In *Geoscience and Remote Sensing Symposium, 2007. IGARSS 2007. IEEE International*, pp.1521–1524, july 2007.
- [17] L. Gomez-Chova, G. Camps-Valls, J. Munoz-Mari, and J. Calpe, "Semisupervised image classification with laplacian support vector machines," *Geoscience and Remote Sensing Letters, IEEE*, Vol.5, pp.336–340, july 2008.
- [18] W. Kim, M. Crawford, and J. Ghosh, "Spatially adapted manifold learning for classification of hyperspectral imagery with insufficient labeled data," In *Geoscience and Remote Sensing Symposium, 2008. IGARSS 2008. IEEE International*, Vol.1, pp.I-213–I-216, july 2008.
- [19] W. Kim and M. Crawford, "Adaptive classification for hyperspectral image data using manifold regularization kernel machines," *Geoscience and Remote Sensing, IEEE Transactions on*, Vol.48, pp.4110–4121, nov. 2010.
- [20] "Multispec© | tutorials," <https://engineering.purdue.edu/~biehl/MultiSpec/hyperspectral.html>. Accessed: 19/01/2012.
- [21] G. Camps-Valls and L. Bruzzone. *Kernel methods for remote sensing data analysis*. Wiley Online Library, 2009.

^۱ multispectral	^{۱۸} Principal Component Analysis	(SVDD)
^۲ hyper-spectral	^{۱۹} Isomap	^{۳۵} Tikhonov
^۳ classification	^{۲۰} residual variance	^{۳۶} Local and General Consistency
^۴ labeling	^{۲۱} Locally Linear Embedding	^{۳۷} regularization
^۵ segmentation	^{۲۲} Local Tangent Space Allignment	^{۳۸} loss function
^۶ curse of dimensionality	^{۲۳} Laplacian Eigenmaps	^{۳۹} Reproducing Kernel Hilbert Space
^۷ feature extraction	^{۲۴} supervised learning	(RKHS)
^۸ change detection	^{۲۵} semi-supervised learning	^{۴۰} hinge loss function
^۹ clustering	^{۲۶} transductive	^{۴۱} Aviris
^{۱۰} data visualizasion	^{۲۷} inductive	^{۴۲} NASA : National Aeronautics and Space
^{۱۱} signal unmixing	^{۲۸} knowledge transfer	Administration
^{۱۲} endmember	^{۲۹} generative	^{۴۳} Indian Pine Teset Site
^{۱۳} model inversion	^{۳۰} Expectation Maximization	^{۴۴} Gaussian Process Learning Method
^{۱۴} segmentation	^{۳۱} Support Vector Machine	^{۴۵} hinge loss function
^{۱۵} unsupervised	^{۳۲} transductive SVM	^{۴۶} embedding space
^{۱۶} one against one	^{۳۳} k-means	^{۴۷} Laplacian Diffusion Kernel
^{۱۷} one against all	^{۳۴} Suppoer Vector Domain Description	^{۴۸} prior