بررسی روشهای طبقه‌بندی تصاویر ماهواره‌ای

دکترمحمداخباری
عضوهیأت علمی دانشگاه امام حسین (ع)

مهندس ابوالفضل رنجبر
کارشناس‌ارشد فتوگرامتری و سنجش از دور

مهندس سیدمحمدباقر فاطمی
کارشناس ارشد سیستم اطلاعات جغرافیایی
عضو هیأت علمی دانشگاه تبریز

چکیده

استخراج اطلاعات از تصاویر ماهواره‌ای در دهه‌های اخیر رشد چشمگیری داشته و در این راستا الگوریتم‌های مختلفی جهت استخراج اطلاعات از تصاویر ماهواره‌ای معرفی شده‌اند که هرکدام از این روشها مزایا و معایبی دارند.

درکل روشهای طبقه بندی تصاویر ماهواره‌ای به دو دسته طبقه بندی نظارت شده و نظارت نشده تقسیم می‌شوند. همچنین روش طبقه بندی نظارت شده خود به دو روش پارامتریک و غیر پارامتریک تقسیم می‌شوند. در این مقاله هدف معرفی و بررسی الگوریتم‌های روشهای طبقه بندی تصاویر ماهواره‌ای نظارت شده از نوع پارامتریک و نظارت نشده‌ها از لحاظ دقت و روش استخراج اطلاعات آنها می‌باشد. درنهایت با مقایسه الگوریتم‌های روشهای موجود به این نتیجه می‌رسیم، روش بیشترین شباهت نسبت به روشهای کمترین فاصله و متوازی السطوح دقت بیشتری دارد اما بازبا این روش دقت مورد نظر در طبقه بندی را نمی توان بدست آورد. در حقیقت از روشهای آماری نظیر بیشترین شباهت زمانی که هدف رسیدن به دقت بالا مورد نیاز باشد، نمی‌توان استفاده کرد. زیرا روش بیشترین شباهت یک روش کاملاً آماریست و بنابراین توانایی پذیرفتن و بکارگیری اطلاعات خارجی در شکل غیر آماری نظیر هندسه تصویربرداری، هندسه عوارض مورد تصویربرداری و همچنین موارد مؤثری چون اتمسفر در فرایند طبقه بندی را ندارد و این یکی از ضعفهای طبقه بندی بیشترین شباهت نسبت به روشهای مدل - پایه محسوب می‌شود. پس جهت بهبود و رفع ایراداتی که در این روش وجود دارد روشهای نظیر مدل - پایه و دانش - پایه معرفی شده‌اند.

لغات کلیدی: طبقه بندی تصاویر ماهواره‌ای، طبقه بندی نظارت شده، طبقه بندی نظارت نشده، کمترین فاصله، متوازی السطوح، بیشترین شباهت، دانش - پایه، مدل - پایه.

۱ - مقدمه

روشهای طبقه بندی تصاویر ماهواره‌ای به دو گروه عمده طبقه بندی نظارت شده و طبقه بندی نظارت نشده تقسیم می‌شوند که در روش طبقه بندی نظارت شده مستقیماً می‌توانیم نقشه موضوعی تهیه کرده و مشکل تفسیر اطلاعات بدست آمده را نداریم مانند تهیه نقشه موضوعی کاربری اراضی. اما این روش هزینه عملیات صحرایی دارد و در واقع برای تهیه نقشه‌های موضوعی نیاز به نمونه‌های آموزشی (نظیر نقشه‌های موجود، عکسهای هوایی، نقاط کنترل زمینی، GIS و غیره) داریم. روشهای طبقه بندی نظارت شده خود به دو گروه اصلی پارامتریک و غیرپارامتریک تقسیم می‌شوند. در روش غیرپارامتریک هیچ فرضی در مورد مدل توزیع داده‌ها نداریم و در واقع هیچ پارامتر آماری در فرایند طبقه بندی محاسبه نمی‌شود. اما در روشهای طبقه بندی پارامتریک، پارامترهای آماری نظیر میانگین و ماتریس واریانس - کوواریانس با استفاده از داده‌های آموزشی برای هر کلاس محاسبه و در فرایند طبقه بندی از آنها استفاده می‌شود. ابتدا به معرفی روشهای ذکر شده پرداخته و سپس، مزایا و معایب آنها ذکر و درنهایت بهترین روش طبقه بندی معرفی خواهدشد. [Richards,1993]

۲ - روشهای طبقه بندی نظارت شده

روشهای طبقه بندی نظارت شده به دو دسته کلی، روشهای پارامتریک و غیرپارامتریک طبقه بندی می‌شوند. از روشهای پارامتریک می‌توان روشهای کمترین فاصله، متوازی السطوح و بیشترین شباهت را نام برد. همچنین از روشهای غیرپارامتریک روشهایی نظیر k- نزدیکترین همسایه، شبکه‌های عصبی و غیره را می‌توان نام برد.

۲-۱ - روشهای پارامتریک

۲-۱-۱ - طبقه بندی به روش کمترین فاصله[1]

در این روش پیکسل مجهول به کلاسی واگذار می‌شود که بردار میانگین آن کلاس نزدیکترین فاصله (فاصله‌ای که در فضای متریک استفاده می‌شود فاصله اقلیدسی است) از پیکسل موردنظر را داشته باشد و یا بیان ریاضی آن:

$$d(x,m_i)=\sqrt{(x-x_{mi})^2+(y-y_{mi})^2} \text{ if } d(x,m_i)<d(x,m_j) \text{ then } X\in i$$

طبقه بندی دیگری نظیر روش کمترین فاصله وجوددارد به اسم Mahalanobis که در واقع در آن اساس باز هم کمترین فاصله می‌باشد اما نه کمترین فاصله اقلیدسی بلکه از کمترین فاصله Mahalanobis استفاده می‌کند که در این روش فاصله به صورت زیر تعریف می‌شود:(در واقع همبستگی بین باندها را هم درنظر می‌گیرد)

$$d(d,m_i)^2=(d,m_i)^t\Sigma^{-1}(d,m_i)$$

۲-۱-۲ - طبقه بندی به روش متوازی السطوح[2]

در این روش حد بالا (v_i) و پایین (ℓ_i) هر متوازی‌السطوح را معرفی کرده و طبق شرط زیر پیکسلها طبقه بندی می‌شوند. به عبارت دیگر اگر پیکسل x در آن محدوده قرار گرفت به آن کلاس تعلق خواهدگرفت.

$$\text{if } u_i\geq x\geq \ell_i \text{ then } x\in i$$

۲-۱-۳ - طبقه بندی به روش بیشترین شباهت

طبقه بندی بیشترین شباهت یکی از عمومی‌ترین روشهای طبقه بندی محسوب می‌شود. در این روش احتمال اینکه یک پیکسل به تمامی کلاسها تعلق دارد محاسبه شده و آن کلاسی که احتمال بیشترین مقدار را داراست، در واقع پیکسل به آن کلاس تعلق می‌گیرد. به بیان ریاضی اگر تمامی کلاسهای طیفی برای یک تصویر را به (ω_i) نشان دهیم که در آن i تعداد کلاسهای موردنظر می‌باشد و احتمالاتی که کلاسها شامل پیکسل موردنظر می‌شوند را به صورت $P(\omega_i|d)$ نشان دهیم. ماکزیمم احتمال را از مقایسه رابطه زیر بدست می‌آوریم:

$$d\in\omega_i \text{ if } P(\omega_i|d)>P(\omega_j|d) \text{ for all } i\neq j$$

اما احتمال $P(\omega_i|d)$ مجهول می‌باشد و ما باید به روشی آنرا تعیین کنیم. ابتدا ما احتمال $P(d|\omega_i)$ (احتمال اینکه یک پیکسل به کدام کلاس تعلق دارد) را از فرمول زیر بدست می‌آوریم. برای این منظور فرض می‌کنیم که داده‌های ما

دارای توزیع نرمال هستند:

$$P(d|\omega_i)\ (2\pi)^{\frac{-n}{2}}\ |\Sigma_i|^{\frac{-1}{2}}\ \exp\{\frac{-1}{2}(d-m_i)^t\ \Sigma_i^{-1}\ (d|m_i)\}$$

که در آن m_i و Σ_i از داده‌های آموزشی بدست می‌آیند و درنهایت $P(\omega_i|d)$ مطلوب از فرمول زیر بدست می‌آید:

(قانون (Bayesian theorem))

$$P(\omega_i|d)=\frac{P(d|\omega_i).P(\omega_i)}{P(d)}$$

$P(\omega_i)$، احتمال اینکه کلاس i در تصویر اتفاق می‌افتد. معمولاً این احتمال معلوم نیست و در این حالت فرض می‌شود که این احتمالات با هم مساویند. احتمال P(d) برای تمامی کلاسها ثابت می‌باشد. پس در کل این دو احتمال $P(\omega_i)$,P(d) در مقایسه دو احتمال حذف می‌شوند و درنهایت داریم: (Strahler 1980)

$$d\in\omega_i \text{ if } P(\omega_i|d)>P(\omega_j|d) \text{ for all } i\neq j$$

۲-۱-۴ - طبقه بندی مدل - پایه [۳]

در طبقه بندی بصری، تصاویر دربرگیرنده دامنه وسیعی از انواع داده‌ها از قبیل درجات خاکستری، اندازه، شکل و الگو در یک وهله زمانی خاص می‌باشد. این درصورتی است که فهم اطلاعات مکانی بوسیله کامپیوتر یک کار مشکلی می‌باشد [Kamiya,2000] درحقیقت شخصی که تفسیر بصری انجام می‌دهد مقدار زیادی دانش اولیه را به منظور بدست آوردن یک تفسیر رضایت بخش به حساب می‌آورد و به جای تفسیر پیکسل به پیکسل، یک ناحیه را از لحاظ خصوصیات مشابه آنها تفسیر می‌کند و در واقع، نتیجه به تجربه شخص مفسر بستگی دارد و شخص مفسر از الگوریتم‌های آماری در تفسیر خود استفاده نمی‌کند. با توجه به مطالب گفته شده نتایج تکنیکهای پردازش تصویر که فقط از آمار استفاده می‌کنند در مقایسه با نتایج حاصل از یک مفسر عکس ماهر ناقص می‌باشند. به منظور اجتناب از خطاهای نظیر همپوشانی رادیومتریکی کلاسهای موجود، اثر پیکسلهای مجاور در مقدار روشنایی یک پیکسل، نسبت اندازه یک پیکسل به اندازه عوارض مورد تصویربرداری و غیره بیشترین توجهات به استفاده از اطلاعات جانبی موجود به عنوان یکی از بهترین راههای بالابردن دقت طبقه بندی معطوف شده است. درعین حال اطلاعات جانبی از محیط مورد تصویربرداری روز به روز کاملتر شده و در دسترس استفاده کنندگان قرارمی‌گیرد.

عکسبرداریهای هوایی، نقشه برداریهای زمینی و به طورکلی سیستم‌های اطلاعات جغرافیایی، داده‌ها و دانش فنی بسیارمفیدی را دراختیار قرارمی‌دهند. چگونگی استفاده از این داده‌ها در فرایند طبقه بندی خود موضوع بحث جدید و گسترده‌ای در زمینه استخراج اطلاعات از تصاویر ماهواره‌ای می‌باشد که موجب توسعه روشهایی نظیر مدل- پایه و دانش - پایه در این زمینه شده است که در ادامه توضیح بیشتری راجع به آنها ارائه شده است. طبقه بندی مدل- پایه براساس مدل کردن بخش جمع آوری داده در یک سیستم سنجش ازدور شامل بخشهای سنجنده، اتمسفر و اشیاء مورد تصویربرداری شکل می‌گیرد. با ایجاد فرضیه‌های مختلف برای تشخیص عوارض موجود در تصویر برپایه مدل‌های مذکور، می‌توان بهترین مقادیر پارامترهای این فرضیه‌ها را تعیین کرد. این کار با استفاده از احتمالات بدست آمده از یک طبقه بندی مرسوم نظیر طبقه بندی بیشترین شباهت و محاسبه هزینه هر مجموعه پارامتر و نهایتاً انتخاب مجموعه پارامتری که کمترین هزینه را بدست می‌دهد، انجام می‌پذیرد. [Fatemi,2001]

همانطوری که گفته شد روش مورد استفاده تخمین کمترین هزینه نامیده می‌شود که هدف آن به حداقل رساندن خطاها در نتایج آنالیز تصویر است. بطوری که این خطاها بوسیله یک تابع تعریف شده توسط استفاده کننده به یک معیار هزینه تبدیل می‌شوند. در این روش ابتدا احتمالات مربوط به هر کلاس برای تمامی پیکسلها محاسبه و نگهداری می‌شوند(E_i). برای این منظور از احتمالات روش بیشترین شباهت استفاده می‌کنیم. دومین مرحله تولید مدل مناسب سیستم سنجش ازدور شامل عوارض مورد تصویربرداری، اتمسفر و سنجنده می‌باشد. این بار براساس مدل پیشنهادی برای هر کلاس فرضیه‌ای ایجاد می‌شود که آنرا با(H_i) نشان می‌دهیم. لازم به توضیح است که هر مدلی که برای سه جزء مذکور بخش جمع آوری‌درنظرگرفته‌شود‌پارامترهایی‌مخصوص به خود دارد. بنابراین با تغییر مقدار این پارامترها می‌توان فرضیه‌های مختلفی را ایجاد نمود. با تغییر مدلها، فرضیه‌ها و در نتیجه نوع و تعداد پارامترها نیز تغییر خواهد کرد. به طور ویژه اگر بتوان به گونه‌های پارامترهای مدل هندسی اشیاء موجود در تصویر را بیابیم در واقع توانسته‌ایم اشیاء و عوارض تصویر را تشخیص داده و نهایتاً تصویر را طبقه بندی کنیم. حل این مسئله برعهده بخش برآورد فرضیه است. بخش برآورد فرضیه براساس هزینه یک فرضیه عمل می‌کند. با تغییر پارامترهای مدل،(H_i)های گوناگونی تولیدمی‌شود. به ازاء هر مجموعه پارامتر خطایی معادل $\Sigma_i^n \Sigma_j^n E_i H_j$ (N به تعدادکلاسها) مرتکب می‌شویم که آنرا هزینه فرضیه موردنظر می‌نامیم. نهایتاً الگوریتم مجموعه پارامتری را انتخاب می‌کنند که کمترین هزینه را بدست دهد. [Fatemi,2001] از کاربردهای این روش می‌توان به هنگام کردن پایگاه داده‌ها را نام برد. پس این طبقه بندی هم محل مرز و هم برچسب آنرا بدست می‌دهد.

۲-۱-۵ - طبقه بندی دانش- پایه [۴]

در این روش طبقه بندی داده‌های موجود در سیستم اطلاعات جغرافیایی به همراه دانش فنی پیرامون مشخصه‌های طیفی و مکانی عوارض زمینی مورد مطالعه جهت بهبود استخراج اطلاعات بکار می‌روند. در واقع به همراه طبقه بندی از دانش فنی دیگری نظیر نقشه‌های خاک، زمین‌شناسی، نقشه‌های خطوط انتقال برق، شبکه‌های جاده‌ها،نقشه‌های توپوگرافی و غیره استفاده می‌شود. در این روش هدف استخراج مشخصه‌های موضوعی و هندسی عوارض زمینی از داده‌های سنجش ازدور می‌باشد. بطورکلی هدف ما از ادغام داده‌ها عبارتست از افزایش دقت طبقه بندی و نیز افزایش قابلیت اطمینان طبقه بندی می‌باشد.

روشهایی نظیر دانش-پایه در تفسیر داده‌های سنجش از دور به منظور بالابردن دقت روشهای طبقه بندی پیکسل مبنا پیشنهاد شده‌اند. تحقیقات اخیر نشان می‌دهند که طبقه بندی دانش- پایه همراه با اطلاعات رقومی GIS نسبت به روشهای پیکسل مبنا و آماری در استخراج اطلاعات کاربری و پوشش زمینی ارجحیت دارد. [Fatemi,2001]

۲-۲ - روشهای غیرپارامتریک

در روش غیرپارامتریک هیچ فرضی در مورد مدل توزیع داده‌ها نداریم و در واقع هیچ پارامتر آماری در فرایند طبقه بندی محاسبه نمی‌شود. از روشهای غیرپارامتریک برای طبقه بندی نظارت شده شبکه‌های عصبی،[۵] K- نزدیکترین همسایه و غیره می‌باشند که خصوصیت خاص خود را داشته و با بقیه روشهای طبقه بندی تفاوت عمده‌ای دارند که در مقالات بعدی ارائه خواهند شد.

۳ - روشهای طبقه بندی نظارت نشده

اغلب به نمونه‌های آموزشی دسترسی نداریم و مجبور به استفاده از روشهایی برای طبقه بندی می‌شویم که نیاز به داده‌های آموزشی نداشته باشند. برای این منظور روشهای طبقه بندی نظارت نشده توسعه داده شده‌اند. از روشهای طبقه بندی نظارت نشده می‌توان به موارد زیر اشاره کرد.

۳-۱ - الگوریتم ISO Data[۶]

الف) تعریف تعداد کلاسترها و حداکثر واریانس کلاسترها

ب) تعریف شرط توقف کلاسترینگ

$|m_i(n+1)-m_i(n)| \le \varepsilon$ که درآن $m_i(n+1)$ میانگین کلاستر در تکرار n+1، $m_i(n)$ میانگین کلاستر در تکرار n

پ) تعیین تعدادی نقطه به عنوان کاندید برای میانگین کلاسترها

ت) اختصاص پیکسلها به‌نزدیکترین مرکز کلاستر (براساس قانون حداقل فاصله)

$d(x,m_i)=\sqrt{(x-x_{m_i})^2+(y-y_{m_i})^2}$ if $d(x,m_i)<d(x,m_j)$ then $x \in i$

ث) محاسبه میانگین جدید کلاستر

ح) بررسی شرط توقف کلاسترینگ "I" بلی بروبه مرحله خ II) خیربروبه مرحله ت"

خ) توقف

۳-۲ - الگوریتم Single Path Clustering

در این روش ما ابتدا یک سری پیکسلها را به تصادف از تصویر اولیه انتخاب می‌کنیم و یک فاصله بحرانی (dc) برای آنها تعریف می‌کنیم. پیکسلها را براساس این فاصله (dc) به پیکسل مرکزی که انتخاب کرده‌ایم اختصاص می‌دهیم. به عبارتی اگر پیکسلها در فاصله کمتر از (dc) پیکسل مرکزی قرارگیرند به آن کلاستر اختصاص داده می‌شوند. در این روش اگر واریانس را زیاد بگیریم تعداد کلاسترها کم خواهد بود. در این روش نیازی به معرفی تعداد کلاسترها نمی‌باشد اما معیار (dc) به طور غیرمستقیم تعداد کلاسترها را تعیین می‌کند.

۳-۳ - الگوریتم Histo Pick Selection

این الگوریتم مانند الگوریتم ISO Data می‌باشد با این تفاوت که انتخاب کاندید برای کلاسترها از طریق هیستوگرام انجام می‌گیرد. این روش نسبت به روش ISO Data سریعتر انجام می‌شود و علت سریعتر بودن این روش اینست که نقاط کاندید انتخابی کلاسترها به میانگین واقعی کلاسترها نزدیکتر هستند. دو الگوریتم برای انتخاب کاندیدها وجوددارد که اولی روش Broad است و اوجهای عمده را درنظرمی‌گیرد که منجر به تعداد کلاسترهای کم در طبقه بندی می‌شود و دومی روش Fine است که اوجهای کوچک را نیز علاوه بر اوجهای بلند درنظر می‌گیرد. روش دوّم، منجر به تعداد کلاسترهای زیاد در طبقه بندی می‌شود.

۳-۴ - الگوریتم درختی[۷]

الگوریتم درختی نیز یکی از روشهای طبقه بندی نظارت نشده می‌باشد. مهمترین مزیت این روش در این است که در آنها نیاز به معرفی تعداد کلاسها وجود ندارد. این الگوریتم‌ها خود دو بخش جمع شونده[۸] و تقسیم شونده[۹] دارند. در روش اول هر پیکسل را به عنوان گروههای تکی درنظر می‌گیرند. سپس گروههای همسایه را با هم براساس فاصله در فضای چندطیفی ادغام می‌کنند و در روش دوم تصویر را به عنوان یک گروه درنظر می‌گیرند و سپس آنها را به گروههای مختلفی تقسیم می‌کنند. در این روش تعداد خوشه‌ها باید ذکر شود. اما در مفهوم این روش نیازی به معرفی تعداد خوشه‌ها نمی‌باشد.

به عنوان مثال مراحل روشهای گروه بندی جمع شونده عبارتند از:

الف) هر پیکسل به عنوان یک کلاستر درنظر گرفته شود.

ب) فاصله بین کلاسترها محاسبه می‌شود. (فاصله اقلیدسی در فضای چندطیفی)

پ) هردوکلاستری که کمترین فاصله‌در نسبت به یکدیگر دارند باهم ادغام می‌شوند.

ت) در مرحله بعدی کلاسترهایی با هم ادغام می‌شوند که دو عضو آنها کمترین فاصله را نسبت به اعضای کلاسهای دیگر داشته باشد.

ث) این روند تا یک سطح دلخواهی ادامه می‌یابد. (البته از لحاظ تئوری این روش می‌تواند تا جایی پیش رود که فقط یک کلاستر باقی بماند)

۴ - نتیجه گیری

● در روش طبقه بندی نظارت شده نیاز به نمونه‌های آموزشی (نظیر نقشه‌های موجود، عکسهای هوایی، نقاط کنترل زمینی، سیستم اطلاعات جغرافیایی و غیره) داشته و مستقیماً می‌توانیم نقشه موضوعی تهیه نمائیم. بعلاوه، مشکل تفسیر اطلاعات بدست آمده را نداریم. مانند تهیه نقشه موضوعی کاربری اراضی. اما این روش هزینه عملیات صحرایی دارد و این درصورتی است که در روش طبقه بندی نظارت نشده ما باید تعداد کلاسترها و حداکثر واریانس هر کلاس را مشخص کنیم. نکته قابل ذکر در مورد انتخاب تعداد کلاسترها به این صورت می‌باشد که اگر تعداد کلاسترها را زیاد انتخاب کنیم مشکل تفسیر و اگر تعداد کلاسترها را کمتر انتخاب کنیم کلاسترهای مخلوط از چندین کلاس خواهیم داشت.

● مزایای روش کمترین فاصله عبارتند از: ۱ - یک روش سریع نسبت به سایر روشهای طبقه بندی نظارت شده پارامتریک می‌باشد. ۲ - به داده‌های آموزشی بیشتری نسبت به روش بیشترین شباهت نیاز ندارد. چون فقط میانگین را محاسبه می‌کند. ۳ - مستقل از توزیع داده‌های آموزشی می‌باشد. ۴ - این روش تمام پیکسلها را طبقه بندی می‌کند. ۵ - همبستگی بین باندها را درنظر نمی‌گیرد. ۶ - محدودیت این روش موقعی ظاهر می‌شود که پیکسلهایی وجود داشته باشند که به هیچ کدام از کلاسها تعلق نداشته باشند و بالاجبار جزء یک محدوده قرارگیرند و درنتیجه طبقه بندی غلط خواهیم داشت. برای رفع این مشکل می‌توان روش ذکرشده را به صورت زیر اصلاح کرد به این صورت که Threshold گذاری کنیم:

if $d(x,m_i) \le d(x,m_j)$ and $d(x,m_i) \le d_c$ then $x \in i$

● مزیت روش کمترین فاصله در این است که نسبت به سایر روشها، زمان کمتر برای طبقه بندی نیاز دارد.

● از معایب روش متوازی السطوح می‌توان موارد زیر را ذکر نمود: ۱ - تمام پیکسلها را طبقه بندی نمی‌کند. ۲ - اگر متوازی السطوح‌ها همدیگر را قطع کنند تصمیم در مورد این پیکسلهای مشترک سخت خواهدبود. در این روش برای اینکه مشکل دوم ایجاد نشود می‌توان کلاسها را براساس اولویت طبقه بندی کرد. ۳ - تعریف محدوده با استفاده از نمونه‌های آموزشی مشکل می‌باشد. دقت این روش از روش طبقه بندی کمترین فاصله کمتراست.

● روش کمترین فاصله در مقایسه با روش طبقه بندی بیشترین شباهت، به محاسبه کمتری نیاز دارد اما از آنجایی که این روش از ماتریس کوواریانس داده‌ها استفاده نمی‌کند بنابراین دقت روش بیشترین شباهت را نخواهد داشت.

● این روش طبقه بندی عمومی‌ترین روش طبقه بندی محسوب شده اما با توجه به دلایل زیر این روش نیز دقت کمتری نسبت به روشهای دانش پایه دارد که این روشها در قسمتهای بعدی این مقاله بحث شده‌اند.

● از مزایای روش بیشترین شباهت این است که این روش با استفاده از میانگین و ماتریس کوواریانس داده‌های تمرینی نتایج کاملتری را نسبت به روشهای نظیر طبقه بندی کمترین فاصله و طبقه بندی متوازی السطوح از داده‌های تصویر بدست می‌دهد. [Richards,1993]

●در زمینه تعیین الگوریتم‌های مناسب برای تعیین کاربری اراضی مطالعات متعددی توسط متخصصان سنجش ازدور در کشورهای مختلف دنیا به عمل آمده است. از مطالعات انجام شده در این مورد می‌توان به کارهای Wilson(۱۹۹۲)، Gong(۱۹۹۲)و Booth(۱۹۸۹) اشاره کرد. روشهای بکار رفته در این تحقیقها شامل روشهای حداکثر احتمال، روش متوازی السطوح و کمترین فاصله از روشهای طبقه بندی نظارت شده می‌باشند. بررسی روشهای ذکر شده براساس دقت کلی و ضریب کاپا با استفاده از تجزیه و تحلیل ماتریس خطا انجام داده‌اند و درکل به این نتیجه رسیده‌اند که روش بیشترین شباهت در بین این روشهای ذکر شده دارای دقت بیشتری می‌باشد.

●با این حال که روش بیشترین شباهت نسبت به روشهای کمترین فاصله و متوازی السطوح دقت بیشتری دارد اما باز با این روش نمی‌توان به دقت موردنظر در طبقه بندی دست یافت. جهت رفع ایراداتی که در این روش وجوددارد روشهای نظیر مدل -پایه و دانش-پایه معرفی شده‌اند. درحقیقت از روشهای آماری نظیر بیشترین شباهت زمانی که هدف رسیدن به دقت بالا موردنیاز می‌باشد، نمی‌توان استفاده کرد زیرا:

● در این روش پیکسل به کلاسی نسبت داده می‌شود که بیشترین احتمال را دارد. در اولین نگاه می‌توان فهمید که یکی از مهمترین نقائص چنین روشی چشم پوشی از دیگر احتمالات کلاسهای دیگر می‌باشد.

● این روش براساس فرض نرمال بودن داده‌ها شکل گرفته است که این فرض نتایج بدست آمده را غیرقابل اطمینان می‌کند.

● الگوریتم بیشترین شباهت طبقه بندی را فقط برمبنای خصوصیات طیفی یک زمان مشخص بکارمی‌برد.

● در این الگوریتم‌ها نمی‌توان از اطلاعات دانش فنی از قبیل داده‌های مربوط به خاک و غیره به سادگی استفاده کرد.

● این الگوریتم از خصوصیات مکانی نظیر اندازه، شکل، بافت و الگو در فرایند طبقه بندی استفاده نمی‌کند.

● روش بیشترین شباهت یک روش کاملاً آماریست و بنابراین توانایی پذیرفتن و بکارگیری اطلاعات خارجی در شکل غیرآماری نظیر هندسه تصویربرداری، هندسه عوارض مورد تصویربرداری و همچنین موارد مؤثری چون اتمسفر در فرایند طبقه بندی را ندارد و این یکی دیگر از ضعفهای طبقه‌بندی بیشترین شباهت نسبت به روشهای مدل- پایه محسوب می‌شود.

● در مرحله تمرینی یک طبقه بندی بیشترین شباهت، اگر در یک سیستم مختصات که هر مؤلفه آنرا یک باند تصویر تشکیل می‌دهد پیکسلهای کلاس معلوم را ترسیم کنیم خواهیم دید که کلاسهای مختلف کاملاً از هم جدا نبوده و با یکدیگر مخلوط می‌شوند. دلایل بسیاری نظیر وجود رفتارهای مشابه در بعضی از پوششهای زمینی در مقابل نور دریافتی، وجود پیکسلهای مختلط و نقائص مرحله تمرینی را می‌توان ذکرکرد. نتایج عملی نشان می‌دهند که روش بیشترین شباهت به مقدار همپوشانی رادیومتریکی کلاسهای مختلف متأثرمی‌شود. دوسری داده‌ها با همپوشانی رادیومتریکی و غیرهمپوشانی رادیومتریکی به این روش طبقه بندی شده که نشان می‌دهند داده‌هایی که غیرهمپوشانی رادیومتریکی داشته‌اند دقت طبقه بندی آنها بیشتر از آن داده‌هایی است که همپوشانی رادیومتریکی داشته‌اند. اما این درصورتی است که روشهای نظیر شبکه‌های عصبی و مدل- پایه نسبت به این مسئله مقاوم است.

● روش بیشترین شباهت نسبت به توزیع داده‌ها حساسیت بیشتری داشته و دقت کلی طبقه بندی برای حالتی که فرض نرمال بودن داده‌ها درست نیست افت قابل توجهی دارد.این درصورتی است که روش مدل- پایه نسبت به این توزیع داده‌ها حساسیتی ندارد و محل مرز را به درستی تعیین می‌کند.(Fatemi,2001)

منابع و مآخذ

1) Richards John .A.(1993):"Remote Sensing Digital Image Analysis , an Introduction", second edition,Springer-Velarg.

2)Kamiya I.(2000):"Imige Classification by Spatial Shift Invariant Neural Network",ISPRS,XXXIII,Part B7,pp.636-639.

3)Fateme S.B.(2001):"A Guided Study to Model-Based Image Analysis ", Thesis,KNT University,Tehran.

4)Benediktsson J.A.,Swain P.H.,Ersoy O.K.and Hong D. (1990) : " Classification of Very High Dimensional Data Using Neural Networks. " IGARSS,pp.1269-1272.

5)Strahelr Alan H.(1980):"The Use of Prior Probability in Maximum Likelihood Classification of Remotly Sensed Data",Remote Sensing of Environment,10:135-163.

6)Booth D.J. and Oldfield R.B.(1989):"A Comparison of Algorithms in Terms of Speed and Accuracy After the Application of a Classification Post Classification Model Filter",International Journal of Remote Sensing. Vol . 10,No.7,pp.1271-1276.

7)Wilson J.D.(1992):"A Comparison of Procedures for Classifying Remotely Sensed Data Using Simulated Data Sets in Corporating Autocorrelation Between Spectral Respanses",International Journal of Remote Sensing , Vol.13,No.2,pp.365-386.

8)Gong P. And howarth P.J.(1992):"Land Use Classification of SPOT HRV Data Using a Cover Frequency Method",International Journal of Remote Sensing,Vol.13,No.8,pp.1459-1471.

7)Strahler Alan H.(1980):"The Use of Prior Probabilities in Maximum Likelihood Classification of Remote Sensed Data" Remote Sensing of Environment,10:135-163.

پی‌نوشت

1)Minimum Distance

2)Paralepiped

3)Model-Based Classification

4)Knowledge-Based Classification

5)Neural Network

6)Iterative Self Organization Data Analysed Thechnique

7)Hierarchical

8)Agglomerative

9)Divisive