انبوهش زدائی نقشه کلاس خاک با استفاده از روش خوشه بندی میانگین های کا و میانگین های سی فازی

شاهرخ فاتحی [۱]، جهانگرد محمدی [۲]، محمدحسن صالحی [۳]، عزیز مومنی [۴]، نورایر تومانیان [۵]، اعظم جعفری [۶]

استادیار بخش تحقیقات خاک و آب، مرکز تحقیقات و آموزش کشاورزی و منابع طبیعی استان کرمانشاه، سازمان تحقیقات و آموزش ترویج کشاورزی، کرمانشاه، ایران، [۲]استاد گروه خاک‌شناسی، دانشکده کشاورزی، دانشگاه شهرکرد، شهرکرد، ایران ، [۳]استاد گروه خاک‌شناسی، دانشکده کشاورزی، دانشگاه شهرکرد، شهرکرد، ایران ، [۴] دانشیار مؤسسه‌ی تحقیقات خاک وآب، سازمان تحقیقات و آموزش ترویج کشاورزی، کرج، ایران ، [۵] دانشیار بخش تحقیقات خاک و آب، مرکز تحقیقات و آموزش کشاورزی و منابع طبیعی استان اصفهان، سازمان تحقیقات و آموزش ترویج کشاورزی، اصفهان، ایران، [۶]استادیار گروه خاک‌شناسی، دانشکده کشاورزی، دانشگاه شهید باهنر کرمان، کرمان، ایران
E-mail: shahrokh.fatehi@gmail.com

چکیده

انبوهش‌زدایی مکانی به مفهوم پیش‌بینی توزیع مکانی کلاس‌های خاک به صورت منفرد می‌باشد. در اینپژوهش انبوهش‌زدایی مکانی کلاس‌های خاک (زیرگروه‌های خاک) در سطح ۲۴ هزار هکتار از نقشه‌ی یک میلیونم خاک‌های ایران در استان کرمانشاه با استفاده از الگوریتم خوشه‌بندی میانگین‌های کا و خوشه‌بندی میانگین‌های سی فازی انجام شد. نقشه‌های مشتقات مدل رقومی ارتفاع، شاخص‌های حاصل از تصاویر ماهواره‌ای و نقشه ویژگی‌های خاک به عنوان داده‌های ورودی مدل مورد استفاده قرار گرفتند. با استفاده از معیارهای اعتبارسنجی درونی، ۷ خوشه به عنوان تعداد خوشه‌ی بهینه برای روش خوشه‌بندی میانگین-های کا و میانگین‌های سی فازی تعیین شد. صحت کلی نقشه‌های رقومی کلاس خاک تولید شده از روش‌های خوشه‌بندی میانگین‌های کا و خوشه‌بندی میانگین‌های سی فازی به ترتیب ۱۶ و ۲۹ درصد بود. نتایج مذکور حاکی از این است که روش خوشه‌بندی میانگین‌های سی فازی کارایی و دقت بیشتری ازخوشه‌بندی میانگین‌های کا برای پیش بینی زیرگروه های خاک دارد.

واژه های کلیدی: انبوهش زدایی مکانی، خوشه‌بندی میانگین‌های کا، خوشه‌بندی میانگین‌های سی فازی، نقشه خاک

مقدمه

نقشه‌های سنتی خاک، منبع عظیم اطلاعات خاک در هر کشوری هستند و میراث و سرمایه ملی محسوب می‌شوند. بنابراین، امروزه پژوهشگران می‌کوشند با استفاده از روش‌های مختلف انبوهش‌زدایی مکانی[۱] در چهارچوب نقشه برداری رقومی خاک، جزییات مکانی و صحت و دقّت نقشه‌های سنتی خاک را بدون نمونه برداری صحرایی اضافی و با کمک متغیرهای محیطی بهبود ببخشند. به طورکلی، در فرایند انبوهش زدایی، اطلاعات نقشه‌های کلاس خاک از قدرت تفکیک مکانی درشت (مقیاس کارتوگرافی کوچک) به قدرت تفکیک مکانی ریز (مقیاس کارتوگرافی بزرگ) تبدیل می‌شود. کلاس‌های خاک در نقشه‌های حاصل از آن با درجه‌ای از عدم قطعیت و به صورت منفرد نشان داده می‌شود.

در نقشه‌برداری سنتی، خاکشناسان ابتدا یک مدل ذهنی از رابطه‌ی خاک-سیمای اراضی در سرتاسر ناحیه مورد مطالعه از طریق یک کار صحرایی متمرکز تهیه می‌کنند. این مدل بیانگر رابطه میان واحدهای سیمای اراضی و انواع خاک‌های محلی است که توسط شرایط محیطی خاصی تعیین می‌گردد. خاکشناسان واحدهای نقشه‌ی خاک را بر اساس فهمشان از توزیع واحدهای سیمای اراضی با استفاده از تفسیر عکس‌های هوایی ترسیم می‌کنند. بنابراین در مدل خاک-سیمای اراضی، نوع خاک

[1] spatial disaggregation

در واحد نقشه مطابق با نحوه‌ی توزیع سیمای اراضی در منطقه مورد مطالعه می‌باشد. روش‌های داده‌کاوی نظارت نشده مانند الگوریتم خوشه‌بندی میانگین‌های کا، با خوشه‌بندی متغیرهای کمکی، ترکیبات محیطی (خوشه‌هایی) تولید می‌کنند که مشابه واحدهای سیمای اراضی با نوع خاک در ارتباط هستند. در همین زمینه برخی محققین نشان داده‌اند که تجزیه تحلیل خوشه‌ای برای آشکارنمودن الگوی همگنی خاک و برای شناسایی رابطه بین خصوصیات خاک ولندفرم‌ها مفید هست (Young & Hammer , 2000). بنابراین ایده پس این مطالعه، همراهی نوع خاک‌ها و ترکیبات محیطی حاصل از خوشه‌بندی است. در این پژوهش از این دانش برای تولید نقشه‌های خاک به هنگام شده و با قدرت تفکیک بالا در رابطه با متغیرهای محیطی دقیق و تفصیلی که قابل دسترس هستند، استفاده شده است.

برای اولین بار بحث انبوهش زدائی نقشه‌های خاک توسط مک برتنی (۱۹۹۸) مطرح گردید و از ابتدای دو دهه‌ی اخیر محققین مختلف برحسب وجود یا کافی نبودن داده‌های نیم‌رخ خاک و تفصیلی بودن یا نبودن نقشه‌های خاک از روش‌های مختلفی برای انبوهش‌زدائی نقشه‌های کلاس خاک استفاده نموده‌اند. پژوهشگران برای انبوهش‌زدایی مکانی نقشه‌های خاک برحسب اطلاعات موجود در راهنمای نقشه (قت نقشه‌ها) و گزارش‌های همراه آن‌ها، انواع روش‌های داده‌کاوی نظارت نشده و نظارت شده مورد استفاده قرار داده‌اند. در جائی که نقشه‌های خاک، فاقد اطلاعاتی از نیم‌رخ‌های خاک هستند و ترکیب اجزاء واحد نقشه معلوم بوده امّا نسبت آن‌ها مشخص نیست از روش طبقه بندی نظارت نشده[۲]برای انبوهش‌زدائی کلاس‌های خاک استفاده شده است. در این زمینه به روش‌های مختلف خوشه‌بندی مانند میانگین‌های کا[۳]و میانگین‌های سی فازی[۴] که توسط برخی محققین استفاده شده است، می توان اشاره نمود (Bui & Moran, 2001;Yang et al.,2011).

در راستای تولید نقشه‌های خاک با مقیاس بزرگ (قدرت تفکیک مکانی ریزتر) از نقشه های با مقیاس کوچک (قدرت تفکیک مکانی درشت‌تر)، پژوهش انبوهش‌زدایی بخش کوچکی از نقشه یک میلیونم خاک های کشور در استان کرمانشاه (این نقشه فاقد نقاط مشاهداتی و تنها دارای اطلاعات نوع خاک است) برای تولید نقشه های کلاس خاک (در سطح زیرگروه) با قدرت تفکیک ۵۰*۵۰ متر، طی سال های ۱۳۹۲ تا ۱۳۹۴طراحی و اجرا شد.

مواد و روش‌ها

به منظور اجرای این پژوهش، ابتدا محدوده‌ی ۲۴۰۰۰ هکتاری بر روی نقشه‌ی یک میلیونم مشخص گردید و نوع خاک‌ها در این محدوده مشخص شد. محدوده‌ی مورد مطالعه منطبق بر نقشه‌ی یک میلیونیوم خاک‌های ایران، هفت کلاس خاک شامل تیپیک کلسی زرالز، لیتیک کلسی زرپتز، تیپیک هاپلوزرپتز، لیتیک زراورتنتز ، تیپیک کلسی زرپتز کرومیک کلسی زررتز و فلونتیک هاپلوزرپتزرا نشان می‌دهد. سپس مشتقات اولیه و ثانویه مدل رقومی ارتفاع و شاخص‌های حاصل از پردازش تصاویر ماهواره‌ای مانند شاخص رس، شاخص اندازه ذرات، شاخص پوشش گیاهی نرمال شده و نقشه‌های کریجینگ بلوکی درصد شن، سیلت، رس، کربنات کلسیم معادل، کربن آلی و سنگریزه خاک سطحی با اندازه پیکسل ۵۰ متر، عنوان متغیرهای کمّی مورد نظر برای تجزیه تحلیل خوشه بندی در نظر گرفته شدند. سپس عملیات استانداردسازی و تجزیه به مولفه‌های اصلی روی این داده‌ها انجام شد که در نهایت ۳۱ مولفه اصلی به عنوان ورودی مدل‌ها انتخاب شدند.

روش‌های خوشه بندی

هدف اصلی از خوشه‌بندی، گروه‌بندی متغیرها یا نمونه‌های مشابه در یک خوشه و متغیرها یا نمونه‌های غیر مشابه در خوشه‌های دیگر است. خوشه بندی، ابزاری مناسب برای گروه بندی داده‌های چند متغیره غیرساختارمند می‌باشد. مرسوم‌ترین روش‌های خوشه بندی، خوشه‌بندی میانگین‌های کا و خوشه‌بندی میانگین سی فازی می‌باشد که در زیر به شرح آن‌ها پرداخته شده است.

روش خوشه بندی میانگین‌های کا

[2] Unsupervised classification
[3] K means
[4] Fuzzy c means

خوشه‌بندی میانگین‌های کا جز دسته غیرسلسله مراتبی قرار می‌گیرد. الگوریتم عامه پسندی برای خوشه‌بندی پایگاه‌های اطلاعاتی با اندازه بزرگ است. این الگوریتم از ساختار محلی حاکم بر داده‌ها برای تعیین خوشه‌ها استفاده می‌کند. با شناسایی مناطق با تراکم بالای داده‌ها، خوشه‌ها تعیین می‌گردند. به همین منظور، این روش بارها اجرا می‌شود تا تابع هدف یا همان مجموع مربعات خطای کل درون خوشه‌ها ، حداقل شود. خوشه‌بندی میانگین‌های کا، طی مراحل زیر اجرا می شود (Hartigan & Wong , 1979):

۱- ابتدا نقاط اطلاعاتی x_i و تعداد خوشه‌ها k مشخص می‌شود. $i = 1.2. \ldots . n.$

۲- مراحل بعدی می‌تواند به شکل‌های زیر ادامه پیدا کند:

- اختصاص تصادفی اولیه نقاط اطلاعاتی به خوشه‌ها، محاسبه مراکز هر خوشه C_j، j=1.2….k

-تعیین مرکز خوشه ها از پیش C_j، j=1.2….k

۳- فاصله‌ی اقلیدسی هر نقطه‌ی اطلاعاتی نسبت به مرکز خوشه‌ی خودش محاسبه می‌شود سپس مجموع مربعات خطا در درون هر خوشه تعیین می‌گردد. در اینجا، خطا به معنای فاصله هر نمونه تا نزدیک‌ترین مرکز هر خوشه (میانگین داده‌های هر خوشه) است. مربعات همه خطاها (ESS) و مجموع آن‌ها از فرمول زیر محاسبه می‌شود:

$$ESS=J=\sum_{j=1}^{K}\sum_{i=1}^{n}\left\|x_{ij}-C_j\right\|^2 \qquad (۱)$$

در این جا C_j ، مرکز j امین خوشه است و x_{ij} داده‌ی نقطه‌ای در j امین خوشه است.

۴- اختصاص دوباره‌ی هر نقطه‌ی اطلاعاتی به نزدیکترین خوشه و به هنگام نمودن هر مرکز خوشه بعد از هر اختصاص.

۵- تکرار مراحل ۳ و ۴ تا زمانی که اختصاص دوباره‌ی نمونه‌ها به خوشه‌ها صورت نگیرد.

الگوریتم خوشه بندی میانگین های سی فازی

الگوریتم میانگین‌های سی فازی به تفصیل توسط بزدک (۲۰۱۳)، مک برتنی و دگروچر (۱۹۹۲) توصیف شده است. در روش خوشه‌بندی میانگین‌های کا هر نمونه یا متغیر به یک خوشه تعلق دارد و مرز بین خوشه‌ها قطعی است اما در روش میانگین‌های سی فازی هر متغیر یا نمونه به حداقل دو خوشه تعلق پیدا می‌کند. مراکز کلاس‌ها به وسیله‌ی حداقل نمودن تابع هدف زیرتعیین می‌شود (بزدک، ۲۰۱۳):

$$J_m = \sum_{i=1}^{n}\sum_{j=1}^{c} u_{ij}^{m}\left\|x_i - C_j\right\|^2 . 1 \le m < \infty \qquad (۲)$$

در اینجا m هر عدد بزرگتر از ۱، u_{ij} درجه عضویت x_i در خوشه‌ی j ؛ x_i، i امین نمونه؛ C_j ، مرکز j امین خوشه و $\left\|x_i - C_j\right\|$ بیانگر تشابه نمونه‌ها و مرکز هر خوشه می‌باشد. در واقع این مولفه، تابع فاصله بین نمونه‌ها و مراکز خوشه را نشان می‌دهد که می‌تواند فاصله اقلیدسی، منهتان یا ماهالانوبیس باشد. الگوریتم خوشه‌بندی فازی از طریق بهینه سازی مکرر تابع هدف مذکور اجرا می‌شود و درجه عضویت u_{ij} و مرکز خوشه C_j توسط توابع زیر تعیین می‌گردد:

$$u_{ij} = \frac{1}{\sum_{k=1}^{c}\left(\frac{\left\|x_i - C_j\right\|}{\left\|x_i - C_k\right\|}\right)^{\frac{2}{m-1}}} \qquad (۳)$$

$$C_j = \frac{\sum_{i=1}^{n} u_{ij}^{m} . x_i}{\sum_{i=1}^{n} u_{ij}^{m}} \qquad (۴)$$

اینجا ε یک معیار توقف است که مقدار آن بین ۰ و ۱ قرار دارد و k مراحل تکرار است.

اعتبارسنجی درونی الگوریتم‌های خوشه‌بندی

برای شناسایی تعداد خوشه‌هایی که به بهترین وجه ساختار مجموعه داده‌ها را توصیف کنند معمولاً از آگاهی و دانش کاربران در مورد داده‌ها استفاده می‌شود امّا، برای بیشترمجموعه داده‌ها، تعداد خوشه بهینه معلوم نیست. در این حالت خوشه بندی برای دامنه‌ای از خوشه‌ها اجرا می‌شود و سپس برای تعیین مناسب‌ترین تعداد خوشه، از برخی شاخص‌ها استفاده می‌گردد.

ارزیابی درونی خوشه‌بندی میانگین‌های کا

شاخص‌های زیادی برای ارزیابی درونی خوشه‌ها وجود دارد که در پژوهش حاضر ازشاخص هارتیگان (۱۹۷۹) استفاده شده است، اساس این شاخص بر جمع مربعات درون خوشه و جمع مربعات بین خوشه‌ها استوار است. در این پژوهش برای اجرای الگوریتم خوشه‌بندی میانگین‌های کا و اعتبارسنجی آن از بسته نرم افزاریcclust در محیط R استفاده شد.

ارزیابی درونی خوشه‌بندی میانگین‌های سی فازی

به منظور اعتبارسنجی درونی روش خوشه بندی میانگین سی فازی از شاخص فوکویاما سوجنو (۱۹۸۹) استفاده گردید. در این پژوهش خوشه‌بندی میانگین‌های سی فازی و اعتبارسنجی آن با استفاده از بسته نرم‌افزاریe1071 در محیط R انجام شد

اعتبارسنجی بیرونی نتایج خوشه بندی

در این پژوهش برای اعتبارسنجی بیرونی نتایج خوشه بندی، رابطه خوشه‌های حاصل از فرایند خوشه‌بندی با مشاهدات نقطه‌ای کلاس‌های خاک مورد بررسی قرارگرفت. برای ارزیابی خطاهای همراه با کلاس‌های منفرد خاک از تست چندجمله‌ای استفاده می‌شود. برای این منظور، چندین معیار کیفی وجود دارد که همگی مبتنی بر ماتریس درهمی هستند. مهمترین این معیارهای کیفی، صحت کلی (Overall accuracy) است

نتایج و بحث

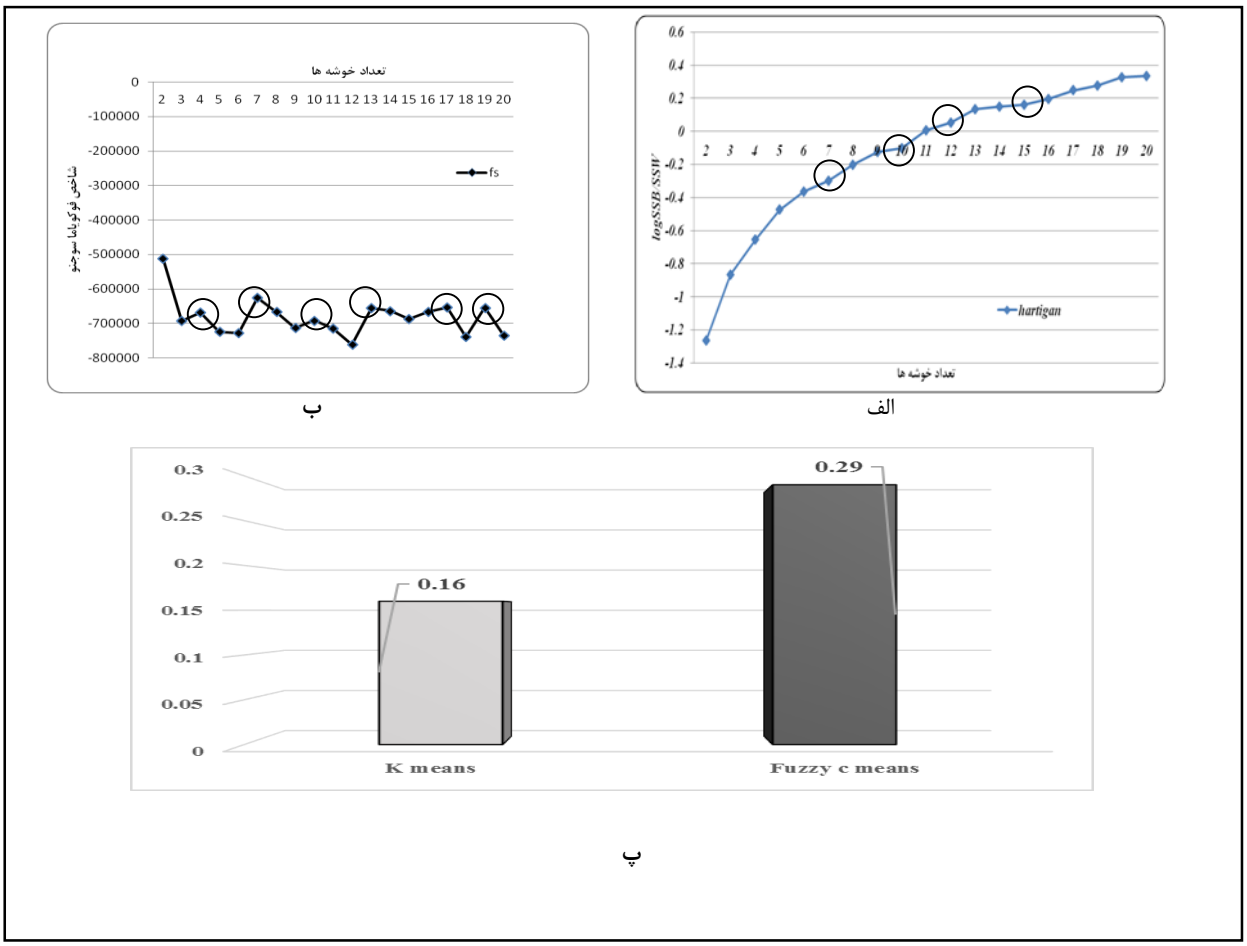
همانطورکه ذکر گردید هدف از خوشه بندی، بازسازی رابطه سیمای اراضی و نوع خاک است. فرض می‌شود متغیرهای محیطی با درجه تشابه بیشتر در یک خوشه قرار می‌گیرند نماینده کلاس خاصی از خاک خواهند بود (Yang et al.,2011). نتایج تهیه نقشه های رقومی کلاس خاک با استفاده از دو روش خوشه بندی میانگین‌های کا و میانگین سی فازی ارائه شده است.

الف) نتایج ارزیابی درونی و بیرونی خوشه بندی میانگین های کا

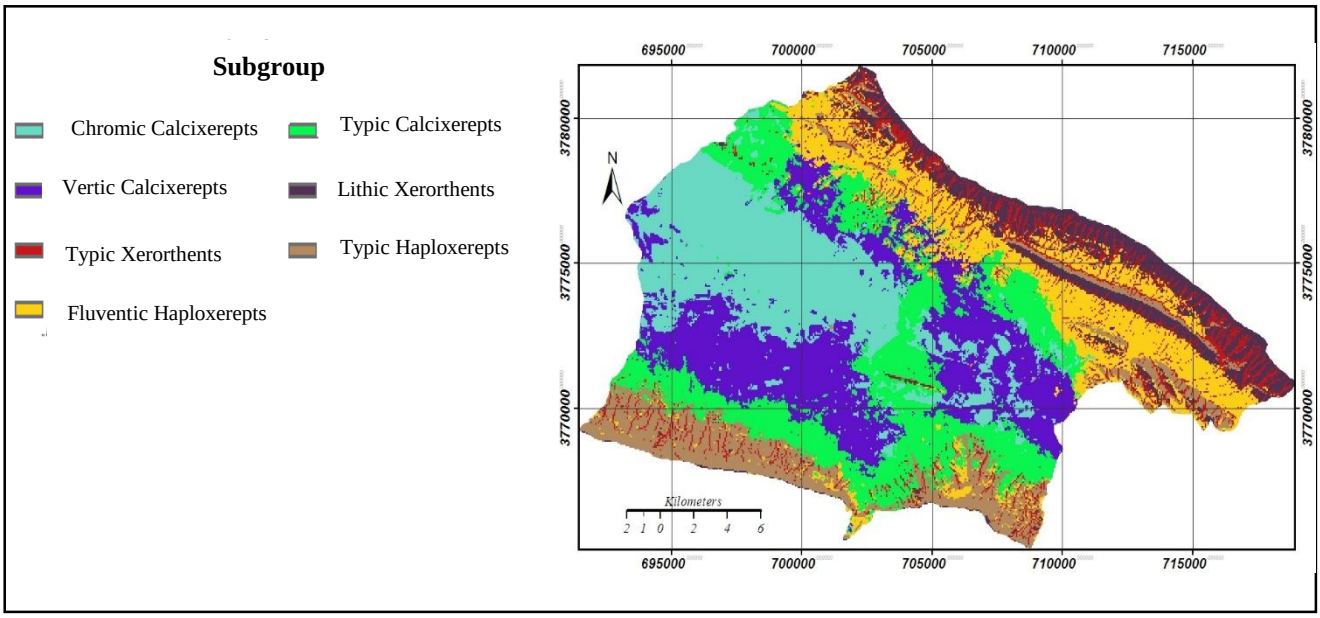
درشکل ۱-الف رابطه‌ی بین شاخص‌های اعتبارسنجی الگوریتم میانگین‌های کا و تعداد خوشه‌ها به صورت منحنی ترسیم شده است. شکست در شیب منحنی در تعداد خوشه برابر با ۷ بیانگر مناسب بودن این تعداد خوشه می‌باشد. سایر تعداد خوشه مناسب قابل ذکر با این روش ۱۰، ۱۲ و۱۵ است. در منطقه مورد مطالعه هفت کلاس خاک (زیرگروه) وجود دارد بنابراین تعداد خوشه ۷ را که با تعداد کلاس‌های خاک (زیرگروه) در منطقه هم خوانی دارد به عنوان تعداد کلاس مناسب انتخاب گردید. بوی و موران (۲۰۰۱) از تعداد کلاس‌های خاک در راهنمای نقشه برای تعیین تعداد بهینه خوشه به عنوان پارامتر ورودی در الگوریتم خوشه بندی میانگین‌های کا استفاده کردند. سپس بر اساس تجربه‌ی صحرایی و رابطه نوع خاک با سیمای اراضی و قرارگیری خوشه‌ها در سیمای اراضی مناسب‌ترین نوع خاک به هر خوشه اختصاص داده شد . سپس نقشه حاصل با استفاده از مشاهدات نقطه‌ای و ماتریس درهمی خطا اعتبارسنجی گردید. نتایج صحت کلی ۱۶ درصد را نشان داد (شکل ۱-پ) که بسیار پایین است.

ب) نتایج ارزیابی الگوریتم خوشه بندی میانگین‌های سی فازی

رابطه‌ی بین شاخص‌های اعتبارسنجی الگوریتم خوشه بندی میانگین‌های سی فازی و تعداد خوشه‌ها به صورت منحنی ترسیم شده‌است (شکل ۱-ب) بهینه بودن تعداد خوشه ۷ در شاخص ارزیابی درونی خوشه بندی فازی محرز است. از آنجا که تعداد کلاس‌های خاک در منطقه مورد مطالعه ۷ زیر گروه خاک است. بنابرین هفت نقشه‌ی عضویت فازی بر اسای فاصله اقلیدسی و درجه فازی بودن ۱/۳ و در ۱۰۰ تکرار تهیه شد. سپس بر اساس حداکثر درجه عضویت، از هفت نقشه درجه عضویت، یک نقشه با هفت خوشه تهیه گردید (شکل ۲). سپس بر اساس دانش تجربی از نحوه قرارگیری خاک‌ها بر روی سیمای اراضی، به هرخوشه یک زیرگروه خاک تعلق گرفت و سپس با استفاده از داده‌های مشاهداتی و ماتریس درهمی خطا صحت این نقشه مورد بررسی قرار گرفت. صحت کلی نقشه تهیه شده ۲۹ درصد است (شکل ۱-پ). به طور کلی این نتایج نشان می‌دهد که روش خوشه‌بندی میانگین‌های سی فازی در پیش بینی و جدا نمودن زیرگروه‌های مختلف خاک دقیق‌تر از روش میانگین‌های کا عمل نموده است



شکل ۱- الف)شاخص *هارتیگان برای* ارزیابی درونی الگوریتم خوشه بندی میانگین‌های کا، شاخص ب) *شاخص فوکویاما سوجنوبرای ارزیابی درونی روش خوشه بندی میانگین‌های سی فازی*،دایره‌ها محل تعداد خوشه مناسب بر روی منحنی را نشان می‌دهند، پ) صحت کلی نتایج خوشه بندی



شکل ۲ نقشه‌ی پیش‌بینی کلاس خاک با روش خوشه‌بندی میانگین سی فازی

منابع

Bezdek J.C. 2013. Pattern recognition with fuzzy objective function algorithms. Springer Science & Business Media. 272p.

Bui E.N.and Moran, C.J. 2001. Disaggregation of polygons of surficial geology and soil maps using spatial modeling and legacy data. Geoderma, 103 (1–2):79–94.

Fukuyama Y. Sugeno M. 1989. A new method of choosing the number of clusters for the fuzzy c-means method. Proceedings of 5th Fuzzy System Symposium. Kobe, Japan.

Hartigan J.A. and Wong M.A. 1979. K-means clustering algorithm. Applied Statistics 28: 100-108.

McBratney A.B. and de Grujiter J.J. 1992. A continuum approach to soil classification by modified fuzzy k-means with extragrades. Journal of Soil Science, 43: 159-175.

McBratney A.B. 1998. Some considerations on methods for spatially aggregating and disaggregating soil information. Nutrient Cycling in Agroecosystems, 50 (1):51–62.

Young F.J. and Hammer R.D. 2000. Defining geographic soil bodies by landscape position soil taxonomy and cluster analysis. Soil Science Society of America Journal, 64: 989-98.

Yang L. Jiao Y. Fahmy S. Zhu A.X. Hann S. Burt J.E. and Qi F. 2011. Updating conventional soil maps through digital soil mapping. Soil Science Society of America Journal, 75: 1044-1053.

Spatial disaggregation soil class map using K means and fuzzy c means clustering

Sh. fatehi[1], J. Mohammadi[2], M. H. Salehi[3], A. Momeni4, N. Toomanian[5] and A. Jafari[6]

[1]Research Assistant Professor, Soil and Water Research Department, Kermanshah Agricultural and Natural Resources Research and Education Center, AREEO, Kermanshah, Iran; E-mail: shahrokh.fatehi@gmail.com

[2]Professor of soil science, college of agriculture, Shahr-e-kord University; E-mail: jahan.mohammad@ymail.com

[3]Professor of soil science, college of agriculture, Shahr-e-kord University; E-mail: Dr.mhsalehi@yahoo.com

[4]Research Associate Professor, Soil and Water Research Institute, AREEO, karaj, Iran; E-mail: momeniham@yahoo.com

[5]Research Associate Professor, Soil and Water Research Department, Esfahan Agricultural and Natural Resources Research and Education Center, AREEO, Esfahan, Iran; E-mail: norairtoomanian@gmail.com

[6]Assistance professor of soil science, college of agriculture, Shahid Bahonar Kerman University; E-mail: a.jafari@uk.ac.ir

Abstract

Spatial disaggregation of soil maps means to predict the spatial distribution of soil classes as individual. In this research, a part of Iran soil map in Kermanshah province at the scale of 1:1000000 was selected with an area of 24000 ha. K means and fuzzy C means clustering were used to disaggregate soil classes (soil subgroups). Landsat images, terrain attributes and map of soil properties were used as input data. Using internal validation, the optimal number of clusters was obtained for fuzzy c means and K means clustering was 7. Overall accuracy of produced soil digital maps of K means and fuzzy c means clustering were 16% and 29%, respectively. This result indicated that the ability of the prediction of the spatial distribution of soil classes using fuzzy c means is better than K means cluster.

Keywords: Spatial disaggregation, K means, fuzzy C means, soil map