بررسی کارایی دو روش C-means و GK برای خوشه‌بندی فازی غلظت مس در اراضی کشاورزی (مطالعه موردی: استان همدان)

مهدی نورزاده حداد[1*]، کاظم خاوازی[2]، محمدجعفر ملکوتی[3]، سید مهدی هاشمی[4]

[*]۱- **نویسنده مسئول:** دانشجوی کارشناسی ارشد، گروه خاکشناسی دانشکده کشاورزی دانشگاه تربیت مدرس(m.nourzade@gmail.com)

۲- استادیار پژوهش موسسه تحقیقات خاک و آب

۳- استاد گروه خاکشناسی دانشکده کشاورزی دانشگاه تربیت مدرس

۴- دانشجوی دکترای سازه‌های آبی دانشکده کشاورزی دانشگاه تربیت مدرس

تاریخ دریافت: ۸۸/۴/۷ تاریخ پذیرش: ۸۹/۸/۲۰

چکیده

مس اگرچه از لحاظ تغذیه گیاه حائز اهمیت می‌باشد؛ ولی یکی از عناصر سنگین بوده و در صورت افزایش زیاد غلظت آن در خاک، به عنوان آلاینده محیط زیست محسوب شده و می تواند برای رشد گیاه و نیز سلامتی انسان مشکلاتی را ایجاد نماید؛ بنابراین بررسی پراکنش غلظت مس در خاک یک منطقه می‌تواند کمک شایانی به مشخص شدن احتمال خطر در سیستم‌های زراعی و در نهایت مدیریت زراعی بنماید. مطالعات قبلی نشان دهنده وابستگی مکانی تغییرات غلظت مس در خاک می‌باشد. روشی که بتواند این وابستگی را در نظر بگیرد، قادر خواهد بود تا تصویر واقعی‌تری از تغییرات خصوصیات خاک را نشان دهد به منظور پهنه بندی متغییر های خاک بعضی از مدل ها نیاز به خوشه بندی یا کلاس بندی هستند و نتایج مدل ها به دقت این تقسیم بندی ها دارد. در این تحقیق از دو روش خوشه‌بندی فازی C-means (Fuzzy c-means clustering) و گاستاسفسن-کسل (GK) به منظور خوشه‌بندی مقدار مس خاک با توجه به pH در اراضی کشاورزی استان همدان استفاده شد. همچنین از پنج شاخص جهت تعیین بهترین روش خوشه‌بندی استفاده گردید. تعداد ۲۱۳ نمونه خاک سطحی (۱۵-۰ سانتی‌متر) از یونجه زارهای استان همدان جمع آوری گردید و غلظت کل مس در آن ها تعیین شد. تعداد بهینه خوشه‌ها بر اساس روش C-means و GK به ترتیب ۸ و ۹ به‌دست آمد. نتایج معیارهای صحت‌سنجی خوشه‌بندی و نیز شکل خوشه‌های ترسیم شده، مشخص کرد که روش GK روش بهتری برای خوشه‌بندی است. با توجه به نتیجه نهایی این تحقیق می‌توان روش GK را برای خوشه‌بندی عناصر سنگین در خاک‌های زراعی پیشنهاد نمود.

کلید واژه‌ها: آلاینده، خاک‌های زراعی، خوشه‌بندی فازی، عناصر سنگین، مس

مقدمه

آلودگی عناصر سنگین در حال حاضر یکی از مشکلات محیط زیست در سطح جهانی می‌باشد (۴). عنصر مس (Cu) نیز یکی از عناصر سنگین محسوب می‌گردد که از منابع مختلف می‌تواند وارد خاک گردد. کشت در اراضی مجاور جاده‌ها و مراکز صنعتی و نیز استفاده زیاد از آفت‌کش‌ها و علف-کش‌ها احتمال افزایش غلظت مس را تا حد مسمومیت گیاه افزایش می‌دهد؛ به طوری که غلظت مس بیش از ۵۰ میلی‌گرم در کیلوگرم خاک در کشور هلند به عنوان غلظت آلوده‌کننده معرفی شده است (۲). با این وجود برخی جوامع گیاهی قادرند به خوبی در خاک‌های غنی از مس رشد نمایند (۵). با توجه به اهمیت غلظت مس به عنوان یکی از عناصر سنگین در اراضی زراعی و امکان ایجاد مسمومیت برای گیاه، پهنه‌بندی غلظت این عنصر به منظور کنترل و مدیریت بهتر آن، در اراضی کشاورزی حائز اهمیت می‌باشد.

به منظور پهنه بندی متغییر های خاک بعضی از مدل ها نیاز به خوشه بندی یا کلاس بندی هستند و نتایج مدل ها به دقت این تقسیم بندی ها دارد روش‌های خوشه‌بندی زیر مجموعه علم داده‌کاوی محسوب می‌شود. داده‌کاوی[1] فرآیند اکتشاف و پردازش پایگاه‌های داده‌ای به منظور استخراج دانش از آنها می‌باشد. به عبارت دیگر داده‌کاوی فرآیند تحلیل مقادیر زیاد داده به منظور اکتشاف الگوهای معنی‌دار و دارای قواعد از میان داده‌ها با استفاده از ابزارهای خودکار و نیمه‌خودکار است (۶). بنابراین تکنیک خوشه‌بندی ابزاری جهت استخراج ساختار اصلی نهفته در مجموعه داده‌ها می‌باشد (۱۳). روش خوشه‌بندی یکی از روش های دسته بندی غیر نظارتی است که در آن یک سری اطلاعات یا داده-ها در قالب یک سری خوشه ها یا گروه ها طبقه بندی می شوند (۱۰). در خوشه بندی غیرنظارتی تعداد خوشه ها از اول مشخص نیست. الگوریتم خوشه‌بندی به دو صورت کلاسیک (قطعی) و فازی قابل انجام است. در شکل کلاسیک خوشه های ایجاد شده به صورت کاملاً جدا و بدون همپوشانی با یکدیگر وجود دارند؛ به طوری که هر شیئی فقط و فقط متعلق به یک کلاس (خوشه) می‌باشد. در خوشه بندی فازی، هر شیئی با درجات عضویت متفاوت، متعلق به تمامی کلاس‌های (خوشه‌های) تفکیکی می باشد، به صورتی که با افزایش مقدار درجه عضویت، مالکیت آن شیئی به آن کلاس بیشتر می‌گردد. خوشه‌بندی فازی در زمینه‌ها و علوم گوناگون از جمله در مسائل مربوط به طبقه‌بندی و پهنه‌بندی عناصر موجود در خاک و آب قابل استفاده می‌باشد. تجریشی و همکاران (۱) با استفاده از تکنیک خوشه‌بندی فازی، پهنه‌بندی آلودگی رودخانه جاجرود را به منظور کنترل، نظارت و پیش بینی تغییرات پارامترهای کیفی رودخانه در اعمال مدیریت کیفیت رودخانه‌ها، تهیه نمودند. شکاری و باقرنژاد (۳) با بررسی کاربرد خوشه‌بندی فازی در طبقه‌بندی خاک‌ها به این نتیجه رسیدند که با وجود این که خاک‌های منطقه مورد مطالعه آنها دارای تمایز کمی بودند؛ ولی روش مزبور، این امکان را فراهم می‌آورد تا این خاک‌ها را تا سطح پدون طبقه‌بندی نمایند. چانگ و چن[2](۹) به منظور بررسی میزان آلاینده‌های خاکی در رسوبات اقیانوسی، تجزیه و تحلیل‌های خوشه‌بندی کلاسیک K-means و خوشه‌بندی فازی را بر مجموعه داده‌ها اعمال نمودند. نتایج تحقیق حاکی از آن بود که خوشه‌بندی فازی به علت وجود مرزهای غیر قطعی بین خوشه‌ها و همپوشانی بین دسته‌های ایجاد شده، نتایج قابل قبولی را ارائه داده است. آنها با استفاده از نتایج خوشه‌بندی مکانی به‌دست آمده و تطبیق آنها با نقشه ساحلی منطقه، توانستند نقشه پهنه‌بندی آلودگی ساحل مورد مطالعه را به‌دست آورند. هدف از انجام این تحقیق اولا، مقایسه و ارزیابی دو روش فازی C-Means و GK[3]، به منظور خوشه‌بندی داده‌های مربوط به غلظت عنصر مس با توجه به pH در یونجه‌زارهای استان همدان و ثانیا با تعیین خوشه‌ها و با توجه به شاخص‌های ارائه شده برای ارزیابی آنها، روش مناسب تعیین گردید.

مواد و روش‌ها

۱- منطقه مورد مطالعه

منطقه مورد مطالعه شامل ۱۲۱۸ کیلومترمربع اراضی کشاورزی استان همدان می‌باشند. این اراضی در موقعیت طول جغرافیائی ′ ۳۴ ° ۴۷ تا ′ ۳۶ ° ۴۹ شرقی و عرض جغرافیائی ′۵۹ °۳۳ تا ′ ۴۸ °۳۵ شمالی قرار دارند. در این تحقیق از ۵۲ مزرعه یونجه نمونه‌برداری صورت گرفت (شکل ۱).

---

1 - Data mining

2- Chang & Cheng

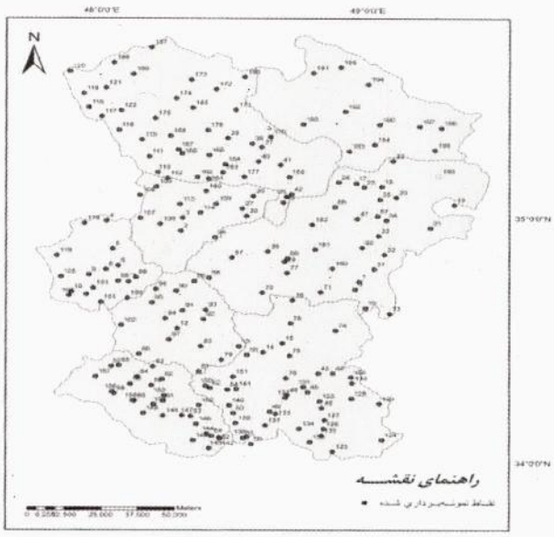
۳- گاستاسفسن-کسل

۲- نقاط نمونه‌برداری شده

نمونه‌های خاک (۱۵-۰ سانتی متر) از ۲۱۳ نقطه به صورت کاملا تصادفی از مزارع یونجه در سطح استان همدان تهیه شدند. مختصات دقیق نقاط نمونه‌برداری به وسیله دستگاه GPS تعیین گردید. در نمونه‌های خاک غلظت کل مس با روش اسید نیتریک ۴ مولار و توسط دستگاه اتمیک اندازه‌گیری شد (۷).

۳- روش‌های خوشه‌بندی فازی

خوشه‌بندی داده‌ها با استفاده از نرم‌افزار MATLAB R2007b انجام گردید. جهت تعیین تعداد بهینه خوشه‌ها، مدل خوشه‌بندی را به تعداد m مرتبه که $2 \le m \le \sqrt{N}$ و N برابر تعداد داده‌ها یا همان ردیف‌های ماتریس داده‌ها است، اجرا کرده و در هر بار اجرای مدل ماتریس مقادیر عضویت و معیارهای صحت‌سنجی محاسبه می‌نمایند. تعداد بهینه خوشه‌ها بر اساس مقدار اکسترمم حاصل شده از هر کدام از معیارهای صحت‌سنجی حاصل می‌شود (۱۲).



شکل ۱- نمایش منطقه مورد مطالعه و پراکنش نقاط نمونه‌برداری شده

۳-۱-الگوریتم خوشه‌بندی فازی C-means (FCM):

تابع هدف الگوریتم خوشه‌بندی FCM برای مجموعه اشیاء X، به صورت زیر تعریف می‌شود:

(۱)

$$J(X;U,V)=\sum_{i=1}^{c}\sum_{k=1}^{N}(\mu_{i,k})^{m}\left\|x_k-v_i\right\|_A^2$$

که در آن $U=[\mu_{i,k}]$ ماتریس تقسیم‌بندی فازی مجموعه X می‌باشد و $m\in(1,\infty)$ توان وزنی است که میزان فازی بودن خوشه‌های نتیجه‌شده را بیان خواهد کرد. $V=[v_1,v_2,\ldots,v_c], v_i\in R^n$ ماتریس مرکز خوشه‌ها[1]، می‌باشد که بایستی در هر مرحله محاسبه شود و $x_k^{(i)}$، شی kام از خوشهٔ i می‌باشد. v_i مرکز جرم خوشهٔ i به شمار می‌رود که از رابطه ۲ به‌دست می‌آید:

$$v_i=\frac{\sum_{k=1}^{N_i}x_k}{N_i}, x_k\in A_i \qquad (۲)$$

در این رابطه N_i تعداد اعضای مجموعه A_i و A_i مجموعه اعضای خوشه iام می‌باشد و برابر مربع حاصل ضرب داخلی معیار فاصله می‌باشد (رابطه ۳).

(۳)

$$\left\|x_k-v_i\right\|_A^2=D_{i,kA}^2=(x_k-v_i)^T A(x_k-v_i)$$

که در آن A اندازه فاصله خواهد بود که بر اساس معیار فاصله اقلیدسی محاسبه می‌گردد (۱۱).

۳- ۲- الگوریتم خوشه‌بندی فازی GK

گاستاسفسن و کسل، الگوریتم استاندارد فازی C-means را با اعمال معیار فاصله تطابق‌پذیر[2]، به منظورجستجوی خوشه‌ها با اشکال متفاوت در یک مجموعه داده‌ها گسترش دادند (۱۱). هر خوشه،

1- Cluster Prototypes or Cluster Centers
2 - Adaptive Distance Norm

ماتریس A_i مربوط به خود را دارد و در نتیجه معیار حاصل‌ضرب داخلی به صورت رابطه ۴ در می‌آید:

(۴)

$$D_{ikA}^2=(x_k-v_i)^T A_i(x_k-v_i),\ 1\le i\le c,\ 1\le k\le N$$

ماتریس‌های A_i به عنوان پارامترهای بهینه‌سازی در عملیات C-means مورد استفاده قرار می‌گیرند. بنابراین، به هر خوشه اجازه می‌دهد تا معیار فاصله آن با ساختار مکانی موضعی داده‌ها سازگاری پیدا کند. با در نظر گرفتن $A_i=(A_1,A_2,...,A_c)$ تابع هدف الگوریتم GK به صورت زیر تعریف می‌شود:

$$J(X;U,V,A)=\sum_{i=1}^{c}\sum_{k=1}^{N}(\mu_{i,k})^m D_{ikA}^2 \qquad (۵)$$

در یک حلقه تکرار و تا مادامی که تابع هدف، مقدار حداقل مطلق خود را به‌دست آورد، مقادیر تابع عضویت، که همان درایه‌های ماتریس مشابهت می‌باشند؛ تغییر خواهند کرد.

۳- ۳- معیارهای صحت‌سنجی خوشه‌بندی فازی

معیارهای صحت‌سنجی خوشه‌بندی فازی استفاده شده در این تحقیق عبارتند از:

۳- ۳- ۱- معیار ضریب تقسیم‌بندی[1] (PC)

این معیار، میزان همپوشانی بین خوشه‌ها را معین می‌کند که توسط بزدک[2] به صورت زیر تعریف شده است:

$$PC(c)=\frac{1}{N}\sum_{i=1}^{c}\sum_{k=1}^{N}(\mu_{i,k})^2 \qquad (۶)$$

که در آن $\mu_{i,k}$ درجه عضویت داده kام به خوشه iام و N تعداد خوشه‌ها می‌باشد. تعداد بهینه خوشه‌ها برابر حداکثر مقدار به‌دست آمده از این معیار خواهد بود (۱۱).

1- Partition Coefficient
2- Bezdek

۳- ۳- ۲- معیار گشتاور رده‌بندی[3] (CE)

میزان فازی بودن خوشه‌های تقسیم‌بندی شده به‌وسیله این معیار اندازه‌گیری می‌شود. تعداد بهینه خوشه‌ها مثل روش قبلی، برابر حداکثر عدد به‌دست آمده از این معیار خواهد بود. تابع این معیار به صورت زیر تعریف شده است (۱۱).

$$CE(c)=-\frac{1}{N}\sum_{i=1}^{c}\sum_{k=1}^{N}\mu_{i,k}\ln(\mu_{i,k}) \qquad (۷)$$

۳- ۳- ۳- شاخص تقسیم‌بندی[4] SC

این معیار، برابر نسبت مجموع فشردگی و جداشدگی خوشه‌ها می‌باشد. در واقع معیار SC برابر مجموع مقادیر معیارهای صحت‌سنجی هر خوشه است که با تقسیم بر درجه فازی بودن آن خوشه نزمال شده است.

(۸)

$$SC(c)=\sum_{i=1}^{c}\frac{\sum_{k=1}^{N}(\mu_{i,k})^m\|x_k-v_i\|^2}{\sum_{k=1}^{N}(\mu_{i,k})\sum_{j=1}^{c}\|v_j-v_i\|^2}$$

حداقل مقدار SC بیانگر بهترین حالت تقسیم‌بندی است. علاوه بر آن، معیار SC زمانی که هدف مقایسه دو حالت خوشه‌بندی متفاوت ولی با تعداد خوشه یکسان باشد نیز استفاده می‌گردد (۸).

۳- ۳- ٤- شاخص جداشدگی[5] (S)

بر خلاف شاخص تقسیم‌بندی SC، شاخص جداشدگی برای بررسی میزان صحت و اعتبار تقسیم‌بندی از حداقل فاصله جداشدگی استفاده می‌کند. زمانی که این شاخص به حداقل مقدار خود

3- Classification Entropy
4- Partition Index
5- Separation Index

میل کند، تعداد بهینه خوشه‌ها به‌دست خواهد آمد (۸).

$$S(c)=\frac{\sum_{i=1}^{c}\sum_{k=1}^{N}(\mu_{i,k})^2\|x_k-v_i\|^2}{N\min_{i,j}\|v_j-v_i\|^2} \quad (۹)$$

۳-۳-۵- شاخص زای و بنی[1] (XB)

زای و بنی یک شاخص برای صحت‌سنجی خوشه‌های فازی تعریف کرده‌اند که هدف آن به کمیت در آوردن نسبت مجموع واریانس درون خوشه‌ها و میزان جداشدگی خوشه‌ها می‌باشد. (۱۰)

$$XB(c)=\frac{\sum_{i=1}^{c}\sum_{k=1}^{N}(\mu_{i,k})^m\|x_k-v_i\|^2}{N\min_{i,j}\|x_k-v_i\|^2}$$

تعداد بهینه خوشه‌ها زمانی ایجاد می‌شود که عدد این معیار حداقل گردد (۸). با استفاده از معیارهای صحت‌سنجی ذکر شده، می‌توان تعداد بهینه خوشه‌ها را مشخص نمود. به‌علاوه از معیار SC می‌توان با معلوم بودن تعداد خوشه‌ها، روش خوشه‌بندی فازی مناسب را تعیین نمود.

نتایج

پس از اندازه‌گیری میزان غلظت مس در نمونه‌های خاک مشخص شد که در هیچ نقطه‌ای غلظت این عنصر، بر اساس استاندارد VSBo (۱۳)، در حد یک آلاینده نیست. در این استاندارد غلظت حد آلایندگی برای مس ۴۰ میلی‌گرم در کیلوگرم ذکر شده است. برخی ویژگی‌های آماری نمونه‌های خاک در جدول ۱ آمده است. منظور تعیین خوشه‌ها هر کدام از مدل‌های خوشه‌بندی ۱۶ مرتبه اجرا گردید. برای تعیین تعداد بهینه خوشه‌ها، مطابق شکل ۲ و ۳، مقادیر ۵ شاخص صحت‌سنجی مورد استفاده در این تحقیق به ترتیب برای روش FCM و GK ترسیم شدند. با توجه به شکل ۲، تعداد بهینه خوشه‌ها بر اساس شاخص CE برابر ۸ خوشه و بر اساس شاخص XB برابر ۹ خوشه شده است. شاخص PC در تعداد ۸ خوشه مقدار کمینه خود را به دست آورده؛ ولی روند تغییرات مقدار این شاخص در تعداد ۸ خوشه تغییر شیب معنی‌دار داشته است، پس می‌توان این نقطه را به عنوان تعداد بهینه خوشه‌ها تلقی نمود. دو شاخص S و SC مقادیر یکسان و نزدیک صفر برای تعداد مختلف خوشه‌ها به‌دست آورده‌اند. بنابراین، این دو شاخص تعداد بهینه‌ای برای خوشه‌بندی معرفی نکرده‌اند. بر این اساس می‌توان استدلال نمود که روش خوشه‌بندی FCM، ۲۱۳ نقطه نمونه‌برداری را در ۸ خوشه جای داده است.

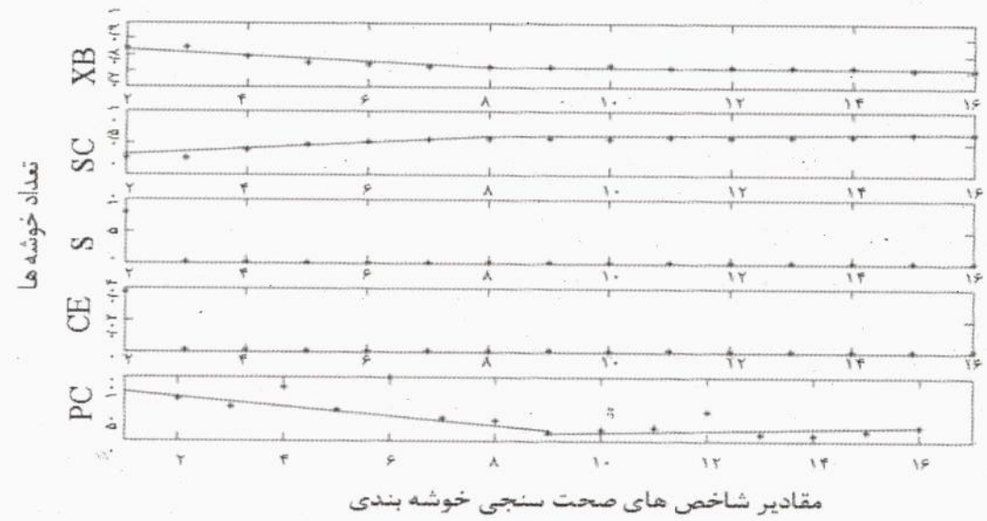
شکل ۳ مقادیر ۵ شاخص مذکور را که از خوشه‌بندی به روش فازی GK به دست آمده، نشان می‌دهد. روند تغییرات شاخص PC در تعداد ۹ خوشه تغییر شیب معنی‌دار داشته است. دو شاخص CE و XB در تعداد ۹ خوشه تغییر شیب معنی‌دار داشته‌اند و به مقادیر اکسترمم خود رسیده‌اند. در صورتی که دو شاخص SC و S قادر به تعیین مقدار بهینه‌ای برای تعداد خوشه‌ها نبوده‌اند. بر اساس نتایج حاصل‌شده تعداد بهینه خوشه‌ها برای روش GK برابر ۹ خوشه می‌باشد.

برای تعیین الگوریتم خوشه‌بندی مناسب از مقادیر شاخص SC استفاده می‌گردد، ولی در مورد مقادیر SC به‌دست آمده در این تحقیق، همان‌طور که از شکل‌های ۲ و ۳ مشاهده می‌شود، مقادیر شاخص مذکور برای هر دو روش FCM و GK تقریباً نزدیک به صفر می‌باشد و نمی‌توان آن‌ها را با هم مقایسه نمود. بنابراین، برای مقایسه دو روش انجام‌شده، خوشه‌های به دست آمده از دو روش خوشه‌بندی فازی در شکل‌های ۴ و ۵ ترسیم

1- Xie & Beni's Index

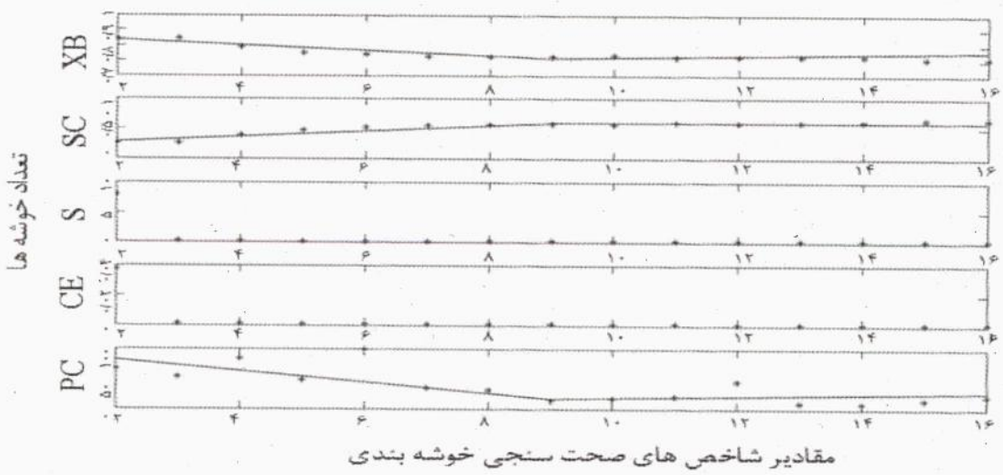
جدول ۱- برخی از شاخص های آماری برای داده‌های غلظت مس (میلی‌گرم بر کیلوگرم)

متغیر	دامنه	کمینه	بیشینه	میانگین	انحراف معیار	واریانس	چولگی	کشیدگی
Cu	۹/۹۶	۰/۲۴	۱۰/۲۰	۱/۲۸	۰/۷۹	۰/۶۲	۶/۸۸	۷/۵۴



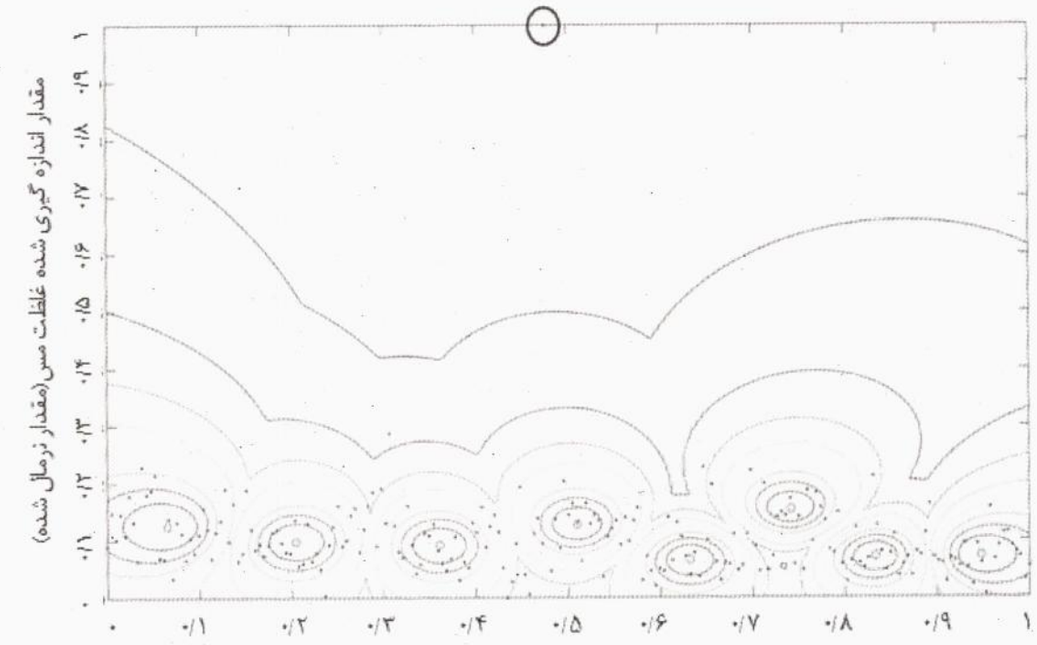
PC: معیار ضریب تقسیم‌بندی، CE: معیار گشتاور رده‌بندی، SC: شاخص تقسیم‌بندی، S: شاخص جداشدگی، XB: شاخص زای و بنی

شکل ۲- مقادیر شاخص‌های صحت‌سنجی در مقابل تعداد خوشه‌ها در روش FCM

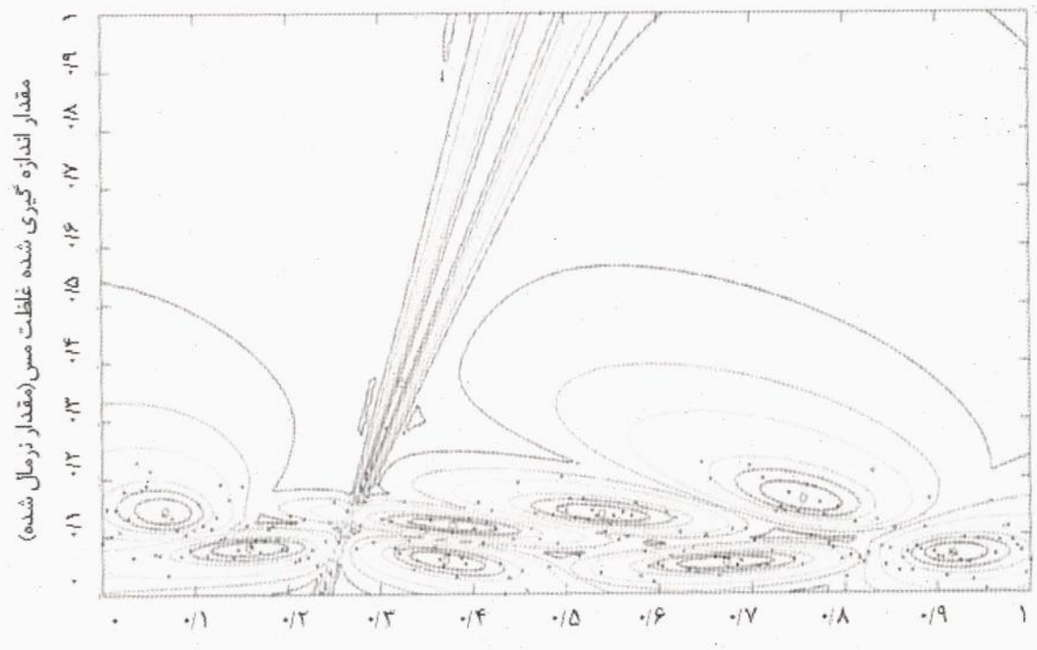


PC: معیار ضریب تقسیم‌بندی، CE: معیار گشتاور رده‌بندی، SC: شاخص تقسیم‌بندی، S: شاخص جداشدگی، XB: شاخص زای و بنی

شکل ۳- مقادیر شاخص‌های صحت‌سنجی در مقابل تعداد خوشه‌ها در روش GK



شکل ٤- خوشه‌های به دست آمده از روش خوشه‌بندی FCM



شکل ٥- خوشه‌های به‌دست آمده از روش خوشه‌بندی GK

گردیدند. در این در شکل ۴، هشت خوشه حاصل شده از روش FCM مشخص شده است. خطوط رنگی اطراف مرکز خوشه‌ها خطوط هم‌درجه عضویت می‌باشد (هر رنگ بیانگر یک درجه عضویت است). خطوط دورتر از هر مرکز خوشه، مقادیر درجه عضویت کمتری را نشان می‌دهند. دو مشکل در خوشه‌بندی حاصل‌شده از روش FCM مشاهده می‌شود؛ اول آن‌که داده مشخص شده با یک دایره مشکی رنگ در شکل ۴ که مقادیر زیادتر غلظت مس نسبت به سایر نقاط هستند، به یک خوشه خاص تعلق نگرفته است. این موضوع سبب می‌شود مقادیر خطوط هم درجه عضویت تحت تاثیر مقادیر بالای این داده‌ها قرار گیرند. مشکل دوم اعمال خوشه‌های تقریباً کروی شکل به تمام خوشه‌ها می‌باشند. با بررسی نتایج دو شکل، محور X (افقی) تعداد خوشه نرمال شده و محور Y (عمودی) مقادیر غلظت مس اندازه‌گیری‌شده هر نمونه می‌باشد. این مقادیر بر اساس روش حداقل-حداکثر نرمال شده‌اند و به مقادیر متناظر در بازه (۰،۱) تبدیل گردیده اند خوشه‌بندی GK، که در شکل ۵ آمده است، گویای آن است که دو اشکال مشاهده‌شده در نتایج خوشه‌بندی FCM رفع شده است.

خوشه‌های ایجاد شده در روش GK دارای شکل‌های بیضی‌شکل هستند که تطابق بیشتری با نحوه توزیع داده‌ها دارند. همان‌طور که در جدول ۲ نشان داده شده است، دو داده با بیشترین مقدار غلظت مس به طور مستقل، در دو خوشه جداگانه ایجاد شده توسط روش GK قرار گرفته‌اند. قرار گرفتن این دو نقطه در دو خوشه جداگانه شاید در نگاه اول منطقی به نظر نرسد؛ ولی این قابلیت بالای الگوریتم خوشه‌بندی GK در تشخیص داده-های پرت می‌باشد. روند کار به این صورت است که داده‌ای که مقدارش اختلاف معنی‌داری با سایر داده‌ها دارد، در یک خوشه مستقل قرار می‌گیرد و از داده های دیگر جدا می‌شود تا مقادیر فواصل سایر داده‌ها از مراکز خوشه‌ها تحت تاثیر این داده قرار نگیرد (جدول ۲). در خوشه‌های ایجاد شده با روش GK، غلظت مس، به صورت یک روند منظم، از خوشه‌های ابتدائی به سمت خوشه‌های انتهائی کاهش می یابد. این در حالی است که در روش FCM، دامنه تغییرات غلظت مس در خوشه‌های ایجاد شده دارای همپوشانی بوده و نمی‌توان روند خاصی را در بین خوشه‌ها مشاهده نمود. بنابراین با توجه به نتایج حاصل شده، روش خوشه‌بندی فازی GK روش مناسب‌تری جهت دسته‌بندی نقاط نمونه‌برداری شده در خوشه‌های مختلف است.

بحث و نتیجه گیری

همان طور که نتایج این تحقیق نشان داد، بر اساس شاخص‌های صحت روش GK کارایی بهتری برای خوشه‌بندی داده‌های مس بر اساس pH داشته است. این در حالی است که در تحقیقات گذشته فقط روش FCM بررسی شده است و به عنوان یک روش مناسب برای خوشه‌بندی ویژگی-های خاک پیشنهاد شده است. در این راستا، امینی و همکاران (۷) با استفاده از روش FCM عناصر سنگین استان اصفهان را خوشه‌بندی نموده و این روش را روشی مناسب برای خوشه‌بندی پیشنهاد نموده‌اند. همچنین شکاری و باقر نژاد (۳) از روش FCM به‌منظور طبقه‌بندی خاک‌ها استفاده کردند و تائید نمودند. این روش کارایی خوبی جهت انجام طبقه‌بندی خاک‌ها داشته است. در این تحقیق نتایج دو روش FCM و GK به‌منظور تعیین روش بهتر با هم مقایسه شدند و با وجود شباهت در نتایج خوشه-ها، مشخص شد که روش GK قابلیت بیشتری در نمایان نمودن اختلافات بین داده‌ها دارد. بر این اساس تعداد خوشه‌ها در روش GK (۹ خوشه) بیشتر از FCM (۸ خوشه) بود. روش‌های خوشه‌بندی در بسیاری موارد براساس تفاوت‌ها ذاتی بین داده‌ها

جدول ۲- توزیع مقادیر عنصر مس در خوشه‌های ایجادشده با دو روش FCM و GK

	FCM			GK	
		غلظت عنصر مس ($mg\ kg^{-1}$)			غلظت عنصر مس ($mg\ kg^{-1}$)
خوشه	تعداد عضو	حداکثر	حداقل	تعداد عضو	حداکثر	حداقل
C1	۱۵	۱/۰۳	۰/۲۴	۱	۱۰/۲۰	۱۰/۲۰
C2	۳۰	۰/۹۹	۰/۴۴	۱	۳/۱۰	۳/۱۰
C3	۲۸	۱/۶۰	۰/۶۸	۶	۲/۵۲	۲/۲۶
C4	۳۳	۰/۹۱	۰/۶۱	۱۷	۲/۲۲	۱/۸۶
C5	۴۴	۱/۷۳	۰/۸۰	۴۰	۱/۸۴	۱/۵۲
C6	۲۷	۲/۲۲	۱/۳۰	۳۳	۱/۵۰	۱/۲۴
C7	۲۵	۲/۴۸	۱/۵۶	۴۰	۱/۲۲	۰/۹۶
C8	۱۰	۱۰/۲۰	۲/۰۵	۵۵	۰/۹۶	۰/۶۶
C9	-	-	-	۲۰	۰/۶۴	۰/۲۴

قادراند تا طبقه‌بندی خوبی از متغیر های مورد نظر ارائه نمایند که در مسائل مدیریت اراضی قابل استفاده است. با توجه به تدریجی بودن تغییرات غلظت عناصر سنگین در خاک و نیز با توجه به ماهیت روش‌های فازی (FCM و GK)، می‌توان گفت روش‌های خوشه‌بندی فازی کارایی بهتری نسبت به روش‌های غیرفازی دارند. همچنین نتایج خوشه‌بندی و کلاس‌های تعیین‌شده را بر اساس توابع عضویت می‌توان به‌صورت نقشه درآورد و به این شکل فهم و نحوه پراکنش خوشه‌ها را در سطح منطقه مورد نظر را به صورت شماتیک بررسی نمود.

در واقع مهم‌ترین کاربرد خوشه‌بندی در این تحقیق، خلاصه سازی منطقه مورد مطالعه مشتمل بر ۲۱۳ نقطه نمونه‌برداری شده به چند خوشه یا به عبارت دیگر به چند منطقه همگن بر اساس مقدار عنصر مس و با توجه به pH خاک در هر نقطه می‌باشد. خلاصه‌سازی به‌خصوص در بحث مدیریت اراضی سبب افزایش قدرت تصمیم‌گیری مدیران شده و با محدود کردن دامنه تصمیم گیری از یک سطح وسیع (اراضی کشاورزی یک استان) به چند خوشه، سبب تسریع در ارزیابی و ارائه راهکارهای مناسب خواهد شد.

منابع

۱. تجریشی، م. ۱۳۸۱. پهنه‌بندی آلودگی رودخانه‌ها توسط تکنیک تحلیل طبقه‌بندی فازی. نشریه ۲۶۱ سازمان مدیریت منابع آب، ۲۶۰ ص.

۲. خداوردی‌لو، ح. ۱۳۸۵. مدل سازی پالایش سبز خاک های آلوده به کادمیم و سرب. پایان‌نامه دکتری خاکشناسی، دانشکده کشاورزی دانشگاه تربیت مدرس، ۱۲۰ ص.

۳. شکاری، پ. و باقر نژاد، م. ۱۳۸۴. بررسی کاربرد روش فازی در طبقه‌بندی خاک‌ها، مطالعه موردی: چشمه‌سفید کرمانشاه، مجله علوم و فنون کشاورزی و منابع طبیعی، شماره ۴، صص ۵۸-۶۸.

۴. علی‌زاده کتک لاهیجانی، ک.، پوری، ح.ع.ر. و امینی، ع.ح. ۱۳۸۴. آلودگی فلزات سنگین در رسوبات دلتای سفید رود. مجله علوم و فنون دریایی ایران، شماره ۳-۴، صص ۴۳-۵۲.

۵. ملکوتی، م.ج.، کشاورز، پ. و کریمیان، ن.ع. ۱۳۸۷. روش‌های جامع تشخیص و توصیه بهینه کودی برای کشاورزی پایدار. انتشارات دانشگاه تربیت مدرس، ۴۵۰ ص.

۶. موسوی، ع. ۱۳۸۶. بخش‌بندی مشتریان خودروهای سواری شهر تهران با استفاده از داده‌کاوی. پایان‌نامه کارشناسی ارشد دانشکده فنی، دانشگاه تربیت مدرس، ۱۱۰ ص.

7. Amini, M., Afyuni, M., Fathianpour, N., Khademi, H. and Fluhler, H., 2005. Continuous soil pollution mapping using fuzzy logic and spatial interpolation. Geoderma 124: 223–233.

8. Balazs, F. 2006. Fuzzy clustering in process of data mining. PhD thesis. Department of Process Engineering. University of Veszprem, Hungry, 157 p.

9. Chang, Y.C., and Cheng, B. 2003. Applying fuzzy cluster method for marine environmental monitoring data analysis. Environmental Informatics Archives, 1; 124-114.

10. Han, J., and Kamber, M. 2006. Data mining concepts and techniques. San Francisco, U.S.A, Morgan Kaufman Publisher, 110 p.

11. Hoppner, F., Klawonn, F., Kruse, R., and Runkler, T. 1999. Fuzzy cluster analysis. Sussex, England: Journal Wiley and Sons, 146 p.

12. Kim, D.W., Lee, K.H., and Lee, D. 2004. On cluster validity index for estimation of the optimal number of fuzzy clusters. Journal of Pattern Recognition Society, 37: 209-225.

13. The Swiss Federal Council, 2008. Regulation on charges of soil 1 July 1998 (as of 1 July 2008) (in Germany) http://www.admin.ch/ch/d/sr/c814_12. html.

14. Valente de Oliveira, J., and Pedrycz, W. 2007. Advances in fuzzy clustering and its applications. Sussex, England, Journal Wiley and Sons, 170 p.