

استخراج پایگاه دانش فازی بهینه به منظور تلفیق و وزن دهی لایه های مکانی در محیط GIS (مطالعه موردی: تهیه نقشه پتانسیل معدنی)

علی اصغر آل شیخ^{۱*}، محمد اصلانی^۲، سعید محسن کلانتری^۳

۱- دانشیار مهندسی GIS، دانشگاه صنعتی خواجه نصیرالدین طوسی، تهران، ایران

۲- دانش آموخته مهندسی GIS، دانشگاه صنعتی خواجه نصیرالدین طوسی، تهران، ایران

۳- استادیار مهندسی زیرساخت دانشگاه ملبورن، استرالیا

پذیرش: ۹۱/۱۰/۶

دریافت: ۹۱/۲/۴

چکیده

وزن دهی و تلفیق لایه های مکان مرجع یکی از مباحث بسیار مهم در آنالیزهای اطلاعات مکانی است. از سوی دیگر، به دلیل عدم قطعیت در لایه های مکانی نیاز به روشی است که بتواند عدم قطعیت را در میزان تأثیر این لایه ها در نظر بگیرد. هدف این تحقیق، استخراج پایگاه دانش بهینه به منظور تلفیق مؤثر لایه های مکانی با در نظر گرفتن عدم قطعیت موجود در آن ها است. در روش های رایج و سنتی، پایگاه دانش براساس تجربه و دانش فرد خبره ایجاد می شود. با وجود این، تبدیل دانش و تجربه فرد خبره به توابع عضویت و قوانین فازی بسیار دشوار است و در اکثر موارد می تواند به نتایج ناکارآمد منجر شود. در این تحقیق دو روش خوشه بندی فازی و الگوریتم ژنتیک برای استخراج پایگاه دانش فازی از مشاهدات محیطی (داده های ورودی - خروجی) به کار رفتند. برای ارزیابی این دو روش، دو پایگاه دانش فازی استخراج شده در یک سیستم استنتاج فازی برای تهیه نقشه پتانسیل معدنی به کار گرفته شدند. نتایج نشان می دهند پایگاه دانش فازی به دست آمده از الگوریتم ژنتیک با RMSE، ۱۶.۲۱ دقت بالاتری را نسبت به خوشه بندی فازی برای داده های آموزشی و کنترل به دست آورده است.

واژه های کلیدی: سیستم استنتاج فازی، الگوریتم ژنتیک، خوشه بندی فازی، نقشه پتانسیل معدنی، سیستم اطلاعات مکانی.



۱- مقدمه

یکی از مهم‌ترین وظایف سیستم‌های اطلاعات مکانی، تحلیل اطلاعات مکان مرجع به‌منظور کشف اطلاعات جدید است (Delaney, 2000: 77). در بسیاری از تحلیل‌های مکانی لازم است نقش و تأثیر پارامترهای مؤثر در پدیده مورد نظر تعیین شود (Hosseinali & Alesheikh, 2008: 1187). بررسی تأثیر معیارها در وقوع یک پدیده مکانی و تلفیق آن‌ها با در نظر گرفتن ساختار و پیچیدگی پدیده مورد نظر مسئله‌ای است که عدم توجه کافی به آن ممکن است به نتایج ناکارآمد و نادرست در تصمیم‌گیری‌های مکانی منجر شود. در دهه‌های اخیر، با پیشرفت سریع علم و فناوری، روش‌های دقیق‌تر و انعطاف‌پذیرتری جایگزین روش‌های سنتی و رایج برای وزن‌دهی و تلفیق لایه‌های مکانی شده است که از جمله آن‌ها می‌توان به شبکه‌های عصبی مصنوعی و سیستم‌های استنتاج فازی اشاره کرد (Leite & Filho, 2008: 675; Porwal, 2006: 179; Tahmasebi & Hezarkhani, 2010: 408). در این روش‌ها، وزن مشخصی به یک معیار نسبت داده نمی‌شود؛ بلکه با توجه به الگوهای مختلف مکانی، تأثیر معیارها متفاوت خواهد بود؛ بنابراین سیستم انعطاف‌پذیری بالایی داشته و به‌شکل هوشمندانه و دقیق‌تری تأثیر معیارها را لحاظ می‌کند (Vahidnia Et al., 2009: 177).

در GIS برخی از موجودیت‌های مکانی که تجزیه و تحلیل می‌شوند، نادقیق و نامعین هستند. منشأ این مسئله را می‌توان در مواردی مانند کمبود دانش، بی‌دقتی در مشاهدات و ماهیت طبیعی خود عارضه دانست (Shad Et al., 2009). سیستم‌های استنتاج فازی از مناسب‌ترین روش‌های محاسباتی برای مدل‌سازی و تحلیل پدیده‌های مکانی با در نظر گرفتن عدم قطعیت هستند. توانایی پیاده‌سازی دانش بشری با استفاده از مفاهیم خاص زبانی و قواعد فازی، غیرخطی بودن، قابلیت تطبیق‌پذیری این سیستم‌ها و دقت بالای آن‌ها در مقایسه با سایر روش‌ها در مدل‌سازی با تعداد محدود داده‌ها، از مهم‌ترین ویژگی‌های این سیستم‌هاست. هسته اصلی سیستم‌های استنتاج فازی پایگاه دانش فازی است. عملکرد بهینه سیستم‌های استنتاج فازی مستلزم طراحی بهینه پایگاه دانش فازی است. در طراحی بهینه سیستم استنتاج فازی دو نکته بسیار حائز اهمیت است: ۱- پیدا کردن تعداد بهینه قوانین و توابع عضویت فازی؛ ۲- فرم

بهینه تواب عضویت فازی (Wong & Her, 1999: 13). در روش های سستی و رایج طراحی پایگاه دانش فازی و تواب عضویت، قوانین فازی براساس دانش و تجربه کارشناس استخراج می شوند. باوجود این، نمی توان دانش کارشناس را به طور کامل در تواب عضویت و قوانین فازی به کار گرفت (Wang & Mendel, 1992: 1414). ضمن اینکه فرایند مصاحبه با کارشناسان و تبدیل دانش و تجربه آنها به تواب عضویت و قوانین فازی بسیار وقت گیر و طاقت فرساست. بنابراین، استخراج خودکار پایگاه دانش فازی با استفاده از مشاهدات محیطی (داده های ورودی - خروجی) راهکاری مناسب برای برطرف کردن این کاستی هاست (Herrera Et al., 1998: 143).

در این تحقیق برای استخراج بهینه تواب عضویت و قواعد فازی از دو روش خوشه بندی فازی و الگوریتم ژنتیک استفاده شده است. همچنین، برای تعیین تعداد بهینه قوانین و تواب عضویت، دو روش خوشه بندی فازی و الگوریتم ژنتیک برای حالت های مختلف تعداد تواب عضویت و قوانین فازی پیاده سازی شد و در نهایت تعداد بهینه پایگاه دانش فازی استخراج شد. در نوشتار حاضر، کوشش شده است سیستم استنتاج فازی و به طور خاص پایگاه دانش فازی برای تلفیق و وزن دهی لایه های مکانی شرح و بسط یابند و در یک کاربرد نمونه ارزیابی شوند. بدین منظور، تهیه نقشه پتانسیل معدنی به عنوان کاربرد مطلوب در نظر گرفته شد. نقشه پتانسیل معدنی می تواند به عنوان مدلی برای تعیین نقاط بهینه حفاری چاه های اکتشافی مورد استفاده قرار گیرد.

۲- مواد و روش ها

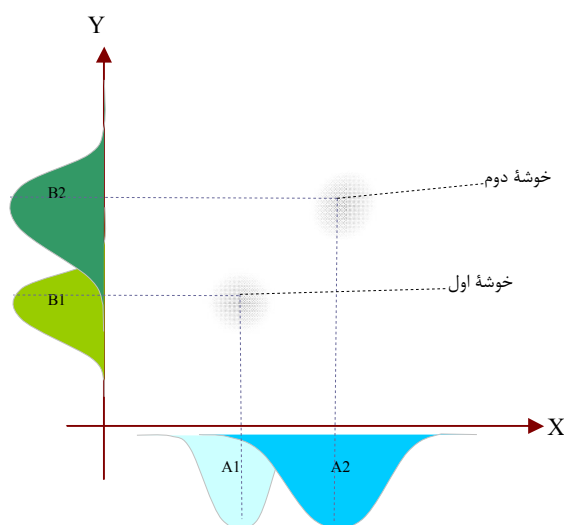
۲-۱- سیستم های استنتاج فازی

سیستم استنتاج فازی از چهار بخش پایگاه دانش فازی (شامل تواب عضویت و قوانین فازی)، فازی ساز، غیرفازی ساز و موتور استنتاج تشکیل شده است. به طور کلی، سیستم استنتاج فازی برپایه استنتاج فازی تشکیل می شود و هسته سیستم یک موتور استنتاج فازی است. موتور استنتاج فازی واحدی است که قواعد و دانش انباشته شده در پایگاه دانش فازی را تحلیل می کند و به نتیجه گیری منطقی منجر می شود. به عبارت بهتر، استنتاج فازی فرایند فرموله کردن یک نگاشت ورودی از داده ها به یک خروجی با استفاده از تواب عضویت و قوانین «اگر -

آنگاه «فازی» است (Siler & Buckley, 2005: 99). یکی از رایج‌ترین موتورهای استنتاج فازی، موتور استنتاج مینیمم است که در این تحقیق استفاده شده است. عملکرد سیستم استنتاج فازی به پایگاه دانش فازی آن وابستگی شدیدی دارد. به عبارت دیگر، اگر پایگاه دانش فازی ناقص و یا حاوی دانشی باشد که مرتبط با دانش مکانی موجود در طبیعت نیست، سیستم استنتاج فازی عملکرد ضعیفی خواهد داشت. بنابراین، استخراج پایگاه دانش فازی بهینه مسئله‌ای بسیار حیاتی است.

۲-۲- خوشه‌بندی فازی برای استخراج پایگاه دانش فازی

با خوشه‌بندی داده‌های ورودی-خروجی (مشاهدات طبیعی) و تصویر کردن خوشه‌ها روی محورهای مختصات می‌توان توابع عضویت و قوانین فازی (پایگاه دانش فازی) را به‌صورت خودکار استخراج کرد. شکل شماره یک نحوه استخراج توابع عضویت و قوانین فازی را برای یک متغیر ورودی و یک متغیر خروجی و دو خوشه نشان می‌دهد.



$$\text{Extracted rules : } \begin{cases} R1: \text{IF } X = A1 \text{ Then } Y = B1 \\ R2: \text{IF } X = A2 \text{ Then } Y = B2 \end{cases}$$

شکل ۱ نمونه‌ای از توابع عضویت و قوانین فازی استخراج‌شده از دو خوشه فازی

براساس شکل شماره یک، تعداد توابع عضویت فازی برای متغیرهای ورودی، خروجی و تعداد قوانین فازی با تعداد خوشه‌ها برابر است (که در شکل ۱ برابر ۲ است). فرض کنید که X^* بردار داده‌های آموزشی با $p+1$ بعد (مؤلفه)، n تعداد داده‌های آموزشی و N تعداد خوشه‌ها باشد. اگر توابع عضویت استخراج شده از نوع گوسین باشند، مرکز تابع عضویت گوسین (C) و پراکندگی آن (σ) با استفاده از روابط زیر تعیین می‌شود:

$$C_j^k = f_j^k \quad \begin{matrix} k = 1, 2, \dots, p+1 \\ j = 1, 2, \dots, N \end{matrix} \quad (1)$$

$$\sigma_j^k = \sum_{i=1}^n \sqrt{\frac{(x_i^{*k} - f_j^k)^2}{2 \log(m_{ij})}} \quad i = 1, 2, \dots, n \quad (2)$$

در روابط ۱ و ۲ C_j^k مرکز تابع عضویت j ام برای بعد k (مؤلفه) f_j^k مرکز خوشه j ام برای بعد k ام، m_{ij} درجه عضویت داده i ام به خوشه j ام، σ_j^k پراکندگی تابع عضویت j ام برای بعد k ام و x_i^{*k} بعد k ام داده آموزشی i ام است. برای خوشه‌بندی فازی روش‌های مختلفی وجود دارد که در این تحقیق از خوشه‌بندی فازی C Means^۱ استفاده شده است (James Et al., 1984: 193).

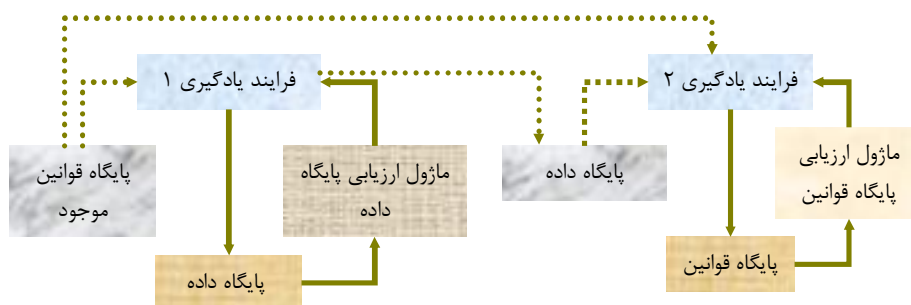
۲-۳- الگوریتم ژنتیک برای استخراج پایگاه دانش فازی

یکی دیگر از راهکارهایی که در این تحقیق برای استخراج پایگاه دانش فازی معرفی می‌شود، استفاده از الگوریتم ژنتیک است. الگوریتم ژنتیک که از روش‌های جست‌وجوی کلی استفاده می‌کند (Alcalá-Fdez Et al., 2009: 906) و توانایی جست‌وجوی مستحکم را فراهم می‌آورد، می‌تواند به‌عنوان ابزار مناسبی برای طراحی و استخراج پایگاه دانش فازی مورد استفاده قرار گیرد (Herrera Et al., 1998: 144). این الگوریتم با مجموعه‌ای از جواب‌ها که

علی اصغر آل شیخ و همکاران استخراج پایگاه دانش فازی بهینه به منظور...

به کروموزوم معروف هستند، کار خود را آغاز می‌کند و سپس مقدار تابع هدف که معیاری برای سنجش کارایی کروموزوم‌هاست، محاسبه می‌شود. حال اگر معیار رسیدن به حالت بهینه برآورده نشده باشد، تولید جمعیت جدید آغاز می‌شود. برای تولید جمعیت جدید از سه روال تکامل: انتخاب، ترکیب و جهش استفاده می‌شود. در ادامه، آن‌قدر الگوریتم تکرار می‌شود که معیارهای پایان الگوریتم باعث توقف آن شوند (Gen & Cheng, 2000: 58-60).

برای استخراج پایگاه دانش فازی با استفاده از الگوریتم ژنتیک روش‌های مختلفی پیشنهاد شده است که ازجمله این روش‌ها می‌توان به میزان‌سازی ژنتیکی پارامترهای پایگاه دانش^۱، یادگیری ژنتیکی قوانین، یادگیری ژنتیکی پایگاه داده و یادگیری ژنتیکی هم‌زمان اجزای پایگاه دانش اشاره کرد (Herrera & Lozano, 2009: 93). در این تحقیق از میزان‌سازی ژنتیکی پارامترهای پایگاه دانش استفاده شده است. شکل شماره دو نحوه عملکرد این روش را نشان می‌دهد.



شکل ۲ میزان‌سازی ژنتیکی پارامترهای پایگاه دانش

(Cordón Et al., 2001)

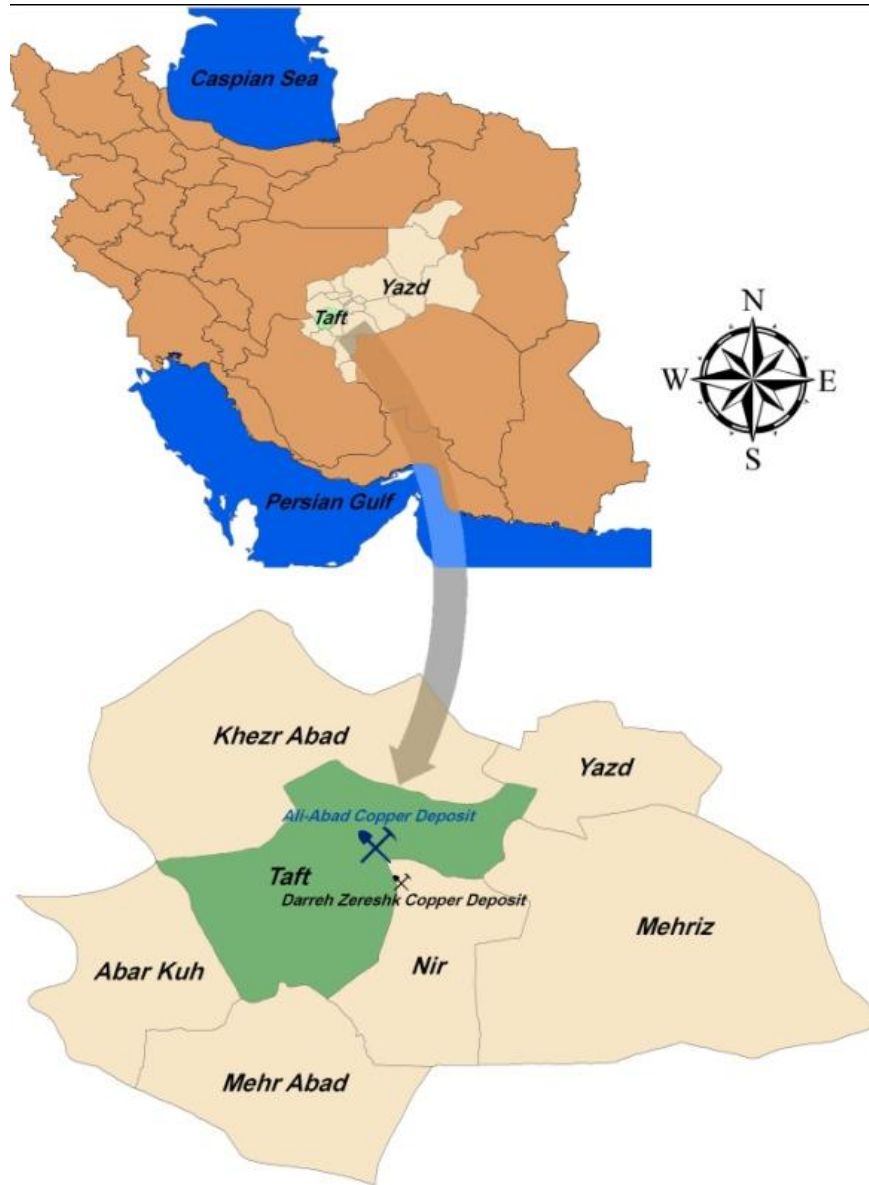
در این روش یک پایگاه داده و یک پایگاه قوانین استنتاج‌شده اولیه وجود دارد و از یک پردازش اولیه برای میزان‌سازی پارامترهای توابع عضویت استفاده می‌شود. نکته حائز اهمیت این است که فرایند میزان‌سازی ژنتیکی فقط فرم توابع عضویت را تنظیم می‌کند و در تعداد ترم‌های زبانی موجود برای هر متغیر تغییری ایجاد نمی‌کند.

1. Genetic Tuning of KB Parameters

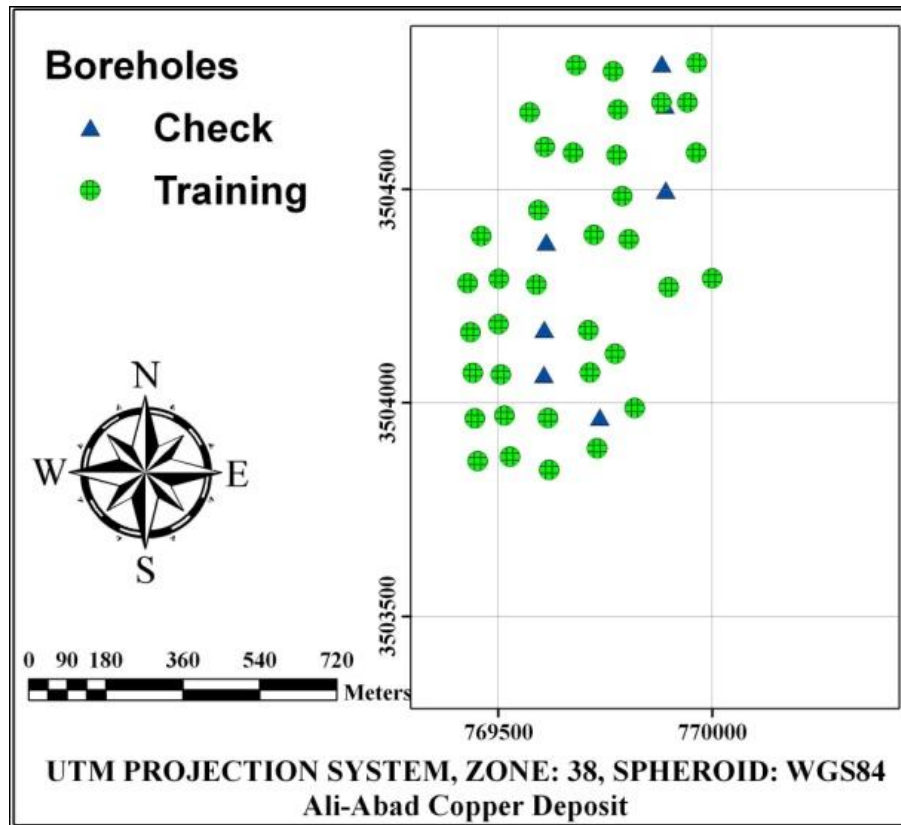
۳- مورد مطالعاتی

برای بررسی نتایج روش های پیشنهادی در این تحقیق، این روش ها برای تهیه نقشه پتانسیل معدنی به کار رفتند. نقشه پتانسیل معدنی به عنوان مدلی از شکل و موقعیت توده های معدنی، دارای اهمیت ویژه ای در مطالعات اکتشاف تفصیلی است و می تواند به انتخاب محل هایی به عنوان هدف برای حفاری به منظور دست یافتن به ذخایر معدنی منجر شود (Tangestani & Moore, 2002: 66). منطقه مورد مطالعه در این تحقیق، کانسار مس علی آباد واقع در شهرستان تفت استان یزد است. کانسار بین دشت کویر مرکزی در شمال شرقی زاگرس و جنوب غرب رشته کوه های شیرکوه قرار دارد. کانسار با مساحت حدود ۱۰۵ کیلومتر مربع در موقعیت جغرافیایی $31^{\circ}39'$ شمالی و $53^{\circ}51'$ شرقی واقع شده است (شکل ۳). از نظر زمین شناسی، کانی سازی در این محل بیشتر در همبندی سنگ های رسوبی، ماسه سنگی، کوارتزیتی با توده های نفوذی آذرین پدید آمده است.

برای تهیه نقشه پتانسیل معدنی به نقشه های معیار تأثیرگذار در تشکیل کانسار مس نیاز است. در این تحقیق برای تهیه نقشه پتانسیل معدنی، فاکتورهای آلتراسیون (شامل فیلک، آرژیلیک و پروپیلیتیک)، سنگ شناسی (شامل دایک، سنگ میزبان و سنگ دیواره)، فاصله از گسل، شدت کل میدان مغناطیسی، شارژ ابیلیته، مقاومت ظاهری و فاکتور فلزی مورد استفاده قرار گرفتند. ویژگی مهم کانسار مس که آن را برای این تحقیق مناسب می سازد، وجود ۴۳ حلقه چاه اکتشافی است. این چاه ها همگی به صورت قائم حفر شده و دارای ژرفایی حدود ۱۰۰ متر هستند و در شبکه ای به نسبت منظم به فواصل تقریبی ۱۰۰ متر از یکدیگر قرار گرفته اند. متوسط عیار عنصر مس نسبت به عمق به عنوان میزان مطلوبیت چاه های اکتشافی در نظر گرفته شد. تمام نقشه های مورد استفاده در این تحقیق از شرکت ملی صنایع مس ایران گرفته شد. مقادیر نقشه های فاکتور در محل های چاه های اکتشافی به عنوان ورودی ها و میزان متوسط درصد خلوص مس به عنوان خروجی بردارهای آموزشی و کنترل در نظر گرفته شدند. همچنین، ۱۵ درصد کل چاه های اکتشافی به عنوان چاه های کنترل (داده های کنترل) و ۸۵ درصد به عنوان چاه های آموزشی (داده های آموزشی) در نظر گرفته شدند. به عبارت دیگر، از ۴۳ چاه اکتشافی ۷ چاه به عنوان کنترل و بقیه آموزشی تعیین شدند (شکل ۴).



شکل ۳ موقعیت کانسار مس علی‌آباد



شکل ۴ موقعیت چاه‌های اکتشافی در منطقه مورد مطالعه

۴- پیاده‌سازی و نتایج

۴-۱- نتایج استخراج پایگاه دانش فازی با استفاده از خوشه‌بندی فازی

در این روش با استفاده از فازی C-Means داده‌های آموزشی خوشه‌بندی شده است و هر خوشه بر روی محورهای مختصات تصویر می‌شود تا پایگاه دانش فازی (توابع عضویت و قوانین فازی) استخراج شود. ورودی اصلی الگوریتم فازی C-Means، تعداد خوشه‌ها (C) است که با تعداد قوانین و دقت مدل‌سازی سیستم استنتاج فازی رابطه نزدیکی دارد. در این تحقیق ۵ سیستم استنتاج فازی مختلف به‌ازای ۶ و ۵ و ۴ و ۳ و ۲ C ایجاد و با استفاده از

چهار معیار خطای برآورد داده‌های آموزشی و کنترل (RSE) (رابطه ۳)، ناسازگاری^۱ و غیرکامل بودن^۲ (Yaochu Et al., 1999: 835) ارزیابی شدند. خطای برآورد داده‌های آموزشی / کنترل با استفاده از رابطه زیر محاسبه می‌شود:

$$RSE = \sqrt{\sum_{i=1}^k (y_{oi} - y_{ci})^2} \quad (3)$$

در رابطه بالا k تعداد داده‌های آموزشی / کنترل، y_{ci} خروجی سیستم استنتاج فازی برای i امین داده (آموزشی / کنترل) و y_{oi} خروجی مورد انتظار است. توابع عضویت از طریق خوشه‌بندی به گونه‌ای تعیین می‌شوند که تا حد ممکن بر داده‌های موجود برازش یابند. از این رو، توزیع توابع عضویت فازی به دست آمده بر مبنای توزیع داده‌های آموزشی است. چنانچه توزیع داده‌ها بی‌قاعده باشد، توابع عضویت فازی غیرواقعی و فاقد معنای فیزیکی خواهند بود. با اعمال دو معیار ناسازگاری و غیرکامل بودن بر پایگاه دانش فازی می‌توان تا حد ممکن از بروز چنین مشکلی جلوگیری کرد. زمانی دو تابع عضویت کامل هستند که اشتراک آن‌ها در حد معقول باشد. همچنین، زمانی قوانین فازی موجود در پایگاه قوانین ناسازگار هستند که قسمت‌های مقدم بسیار شبیه بوده، ولی قسمت‌های تالی متفاوت باشند. به عبارت دیگر، اگر قوانین دارای بخش مقدم یکسان و تالی یکسان باشند، در این صورت درجه سازگاری به یک نزدیک می‌شود. در این مقاله برای محاسبه ناسازگاری و غیرکامل بودن از شاخص شباهت فازی استفاده شده است (Yaochu Et al., 1999: 831).

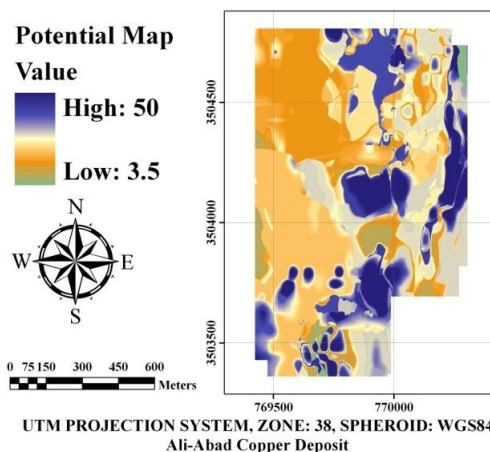
جدول شماره یک مقادیر نرمال شده چهار معیار خطای برآورد داده‌های آموزشی، خطای برآورد داده‌های کنترل، ناسازگاری و غیرکامل بودن و مجموع وزن دار آن‌ها را به‌ازای تعداد مختلف خوشه‌ها نشان می‌دهد.

1. inconsistency
2. incompleteness

جدول ۱ مقادیر نرمال‌شده خطا، ناسازگاری، غیرکامل بودن و مجموع وزن‌دار آن‌ها

۶	۵	۴	۳	۲	تعداد خوشه‌ها (C)
۱	۷۵.۴۶	۶۰.۴۴	۶۲.۳	۱۰۰	خطای برآورد داده‌های آموزشی (خلوص مس، نرمال‌شده)
۹۶.۶۳	۹۲.۰۵	۹۵.۶۶	۱۰۰	۱	خطای برآورد داده‌های کنترل نرمال‌شده (خلوص مس، نرمال‌شده)
۶.۷۱	۲۰.۸۶	۴.۴۲	۱	۱۰۰	ناسازگاری (نرمال‌شده)
۱۰۰	۷۷.۹۰	۵۲.۳۰	۳۰.۴۲	۱	غیرکامل بودن (نرمال‌شده)
۴۰۰	۶۷۶	۵۸۵	۵۸۰	۵۰۴	مجموع وزن‌دار چهار معیار

براساس داده‌های جدول شماره یک، کمترین میزان خطای برآورد داده‌های آموزشی، کنترل، ناسازگاری و غیرکامل بودن به ترتیب به‌ازای ۲ و ۳ و ۶ $C =$ است. بنابراین، با اطمینان کامل نمی‌توان گفت که به‌ازای چه تعداد خوشه، سیستم استنتاج فازی دارای بهترین عملکرد است. از سوی دیگر، بالاترین هدف در طراحی سیستم‌های استنتاج فازی، افزایش هم‌زمان دو معیار دقت (خطای برآورد داده‌های آموزشی و کنترل) و معناپذیری (سازگاری و کامل بودن) است. اما این دو معیار اغلب در جهت عکس یکدیگر عمل می‌کنند. به این معنا که با افزایش بیش از حد معناپذیری، دقت کاهش می‌یابد و با افزایش بیش از حد دقت، معناپذیری کاهش می‌یابد. برای حل این مشکل در این تحقیق از مجموع وزن‌دار این چهار معیار استفاده شد.



شکل ۵ نقشه پتانسیل معدنی زمین‌شناسی با استفاده از خوشه‌بندی فازی



همان طور که در جدول شماره یک دیده می شود، کمترین مجموع وزن دار به ازای $C = 6$ است. بنابراین، پایگاه دانش فازی به دست آمده به ازای ۶ خوشه بهینه ترین پایگاه دانش فازی و سیستم استنتاج فازی ایجاد شده براساس این پایگاه دانش دارای بهترین عملکرد است. با وارد کردن تمام نقشه های فاکتور به سیستم استنتاج فازی به دست آمده نقشه پتانسیل معدنی مس ایجاد شد (شکل ۵).

۴-۲- نتایج استخراج پایگاه دانش فازی با استفاده از الگوریتم ژنتیک

در این روش برای به دست آوردن فرم بهینه توابع عضویت و قوانین فازی موجود در پایگاه دانش فازی از الگوریتم ژنتیک استفاده می شود. ابتدا با در نظر گرفتن یک پایگاه قوانین اولیه فرم بهینه توابع عضویت استخراج شد. سپس با ثابت فرض کردن پایگاه داده استخراج شده، فرم بهینه قوانین فازی استخراج شد. بنابراین، برای شروع بهینه سازی پایگاه دانش فازی تعریف یک پایگاه دانش اولیه ضروری است. در این تحقیق، پایگاه دانش فازی استخراج شده از روش خوشه بندی فازی به عنوان پایگاه دانش اولیه برای الگوریتم ژنتیک در نظر گرفته شد. به دلیل اینکه ۵ پایگاه دانش فازی به ازای ۶ و ۵ و ۴ و ۳ و ۲ $C =$ در بخش ۴-۱ با استفاده از خوشه بندی فازی تعریف شده، بهتر است الگوریتم ژنتیک بر روی تمام این ۵ پایگاه دانش فازی اجرا شود تا بهترین پایگاه دانش انتخاب شود. چون تعداد توابع عضویت و قوانین فازی در تمام پایگاه های دانش فازی یکسان نیستند، طول کروموزوم های نشان دهنده آنها نیز با یکدیگر متفاوت خواهند بود. ذکر این نکته لازم است که توابع عضویت به کار رفته در این روش گوسین هستند و با دو پارامتر (σ و C) در هر کروموزوم قابل کدگذاری اند. پس از کدگذاری کروموزوم ها نوبت به تعریف اپراتورهای ادغام و جهش می رسد. در این تحقیق، روش ادغام دو نقطه با نرخ عملکردی ۰.۷ مورد استفاده قرار گرفت. همچنین نرخ عملگر جهش ۰.۱ انتخاب شد.

خروجی الگوریتم ژنتیک (فرم توابع عضویت و قوانین فازی) به تعریف تابع هدف وابستگی زیادی دارد. دقت برآورد داده های آموزشی یکی از مهم ترین شاخص هایی است

که در تابع هدف الگوریتم ژنتیک در نظر گرفته می شود. اگر فقط این شاخص در تابع هدف لحاظ شود، توابع عضویت و قوانین فازی معنای زبانی خود را تا حدود زیادی از دست می دهند. همچنین اگر شاخص معنادار بودن بدون در نظر گرفتن دقت لحاظ شود، توابع عضویت معنادار خواهد بود؛ ولی دقت برآورد داده های آموزشی به شدت کاهش می یابد. بنابراین، با افزایش دقت، معنادار بودن کاهش می یابد و با افزایش معنادار بودن، دقت کاهش می یابد. بنابر اصل ناسازگاری لطفی زاده^۱ (28: 1973) می توان یک حد آستانه در نظر گرفت که در آن حد آستانه دو شاخص دقت و معنادار بودن در بیشترین مقدار خود قرار دارند و چنانچه یکی از این دو شاخص از این حد آستانه بالاتر رود، شاخص دیگر به شدت کاهش می یابد. بنابراین، باید به گونه ای این دو شاخص در الگوریتم ژنتیک لحاظ شوند که هر دو به بالاترین حد آستانه دست پیدا کنند. برای رسیدن به این هدف، سه شاخص سازگاری، کامل بودن (برای اعمال معناپذیری) و خطا در تابع هدف لحاظ شدند (Yaochu Et al., 1999: 836). شرط توقف نیز ثابت ماندن بهترین مقدار تابع هدف در تعداد نسل های زیاد (در این تحقیق ۲۰) انتخاب شد. دلیل انتخاب چنین شرط توقفی، فقط افزایش ضریب اطمینان از همگرایی الگوریتم ژنتیک به بهینه کلی^۲ است. در گام بعدی با ثابت نگه داشتن فرم بهینه توابع عضویت فازی به استخراج فرم بهینه قوانین فازی پرداخته شد. از تنظیمات ژنتیکی استفاده شده در فرایند مرحله اول نیز در این مرحله استفاده شد. با بررسی قوانین فازی به دست آمده مشخص شد که این قوانین دقیقاً مشابه قوانین استخراج شده از طریق خوشه بندی فازی هستند. این موضوع نشان می دهد قوانین فازی اولیه استخراج شده از طریق خوشه بندی دارای بهترین فرم ممکن بوده اند. جدول شماره دو مقادیر نرمال شده چهار معیار خطای برآورد داده های آموزشی، خطای برآورد داده های کنترل، ناسازگاری و غیر کامل بودن و مجموع وزن دار آن ها را به ازای تعداد مختلف خوشه ها نشان می دهد.

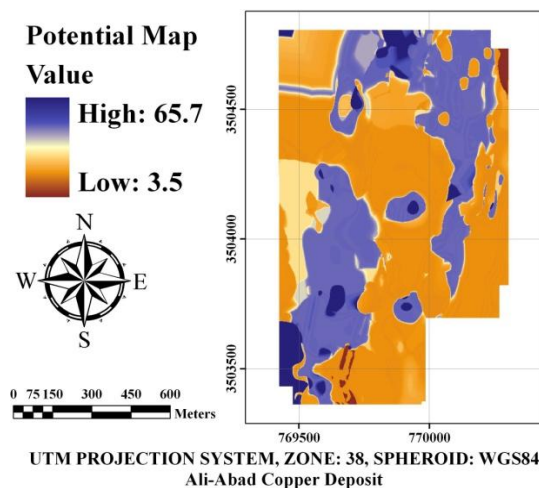
1. L.A. Zadeh

2. global optimum

جدول ۲ مقادیر نرمال شده خطا، ناسازگاری، غیرکامل بودن و مجموع وزن دار آن‌ها

۶	۵	۴	۳	۲	(C) تعداد خوشه‌ها
۱۰۰	۸۹.۹۳۵	۱	۳۸.۰۲۷	۱۹.۱۲۹	خطای برآورد داده‌های آموزشی (خلوص مس، نرمال شده)
۸۸.۹۸۲	۱۰۰	۶۵.۹۶۰	۱	۸۱.۶۳۵	خطای برآورد داده‌های کنترل (خلوص مس، نرمال شده)
۱۰۰	۲۹.۰۱۱	۱.۰۰۰۰۱	۱.۳۸۴	۱	ناسازگاری (نرمال شده)
۱۰۰	۱	۱.۹۲۴	۱.۲۷۶	۲.۹۹۸	غیرکاملی (نرمال شده)
۸۶۶.۹۴۷	۶۸۹.۷۵۰	۲۰۴۸۰۴	۱۵۷.۷۶۹	۳۲۵.۴۱۹	مجموع وزن دار چهار معیار

با توجه به جدول شماره دو کمترین مجموع وزن دار به‌ازای $C = ۳$ است؛ بنابراین پایگاه دانش فازی به‌دست آمده به‌ازای ۳ خوشه بهینه‌ترین پایگاه دانش فازی و سیستم استنتاج فازی ایجاد شده براساس این پایگاه دانش دارای بهترین عملکرد است. با مقایسه دو جدول شماره یک و دو به این نکته پی می‌بریم که الگوریتم ژنتیک دو معیار معنابذیری و دقت را بهتر از روش خوشه‌بندی فازی اعمال کرده است. با وارد کردن تمام نقشه‌های فاکتور به سیستم استنتاج فازی به‌دست آمده نقشه پتانسیل معدنی مس ایجاد شد (شکل ۶).



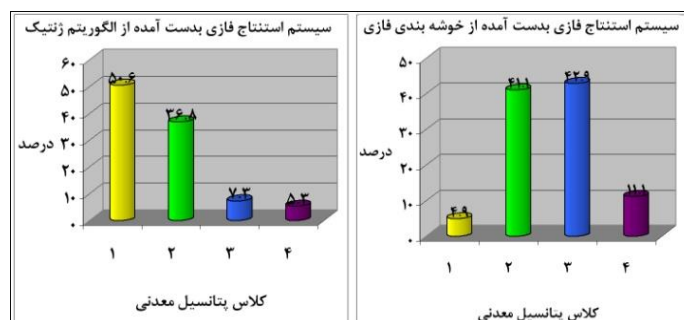
شکل ۶ نقشه پتانسیل معدنی زمین‌شناسی با استفاده از سیستم استنتاج فازی ژنتیک

۵- بحث و نتیجه گیری

برای مدل سازی های مکانی پیچیده به دانشی نیاز است که با استفاده از آن بتوان لایه های مکانی را به طور مؤثر وزن دهی و تلفیق کرد؛ به گونه ای که نتیجه به دست آمده تا حد امکان به واقعیت موجود نزدیک باشد. استفاده از مشاهدات محیطی (داده های ورودی - خروجی) جهت استخراج دانش می تواند راهکاری مناسب در این زمینه باشد. از سوی دیگر، به دلیل وجود عدم قطعیت در اغلب پدیده های مکانی نیاز به طراحی و توسعه روشی است که بتواند عدم قطعیت را در دانش استخراج شده لحاظ کند. در این پژوهش، سیستم های استنتاج فازی به دلیل تکیه بر استنتاج ها و پایگاه دانش فازی به کار رفتند.

با توجه به مطالعه موردی تحقیق حاضر مبنی بر تهیه نقشه پتانسیل معدنی در کانسار مس علی آباد، دو روش خوشه بندی فازی و الگوریتم ژنتیک برای استخراج پایگاه دانش فازی در سیستم استنتاج فازی به کار گرفته شد. برای بررسی عملکرد پایگاه های دانش استخراج شده، لازم است نتایج آن ها در تهیه نقشه پتانسیل معدنی ارزیابی شود.

دو نقشه پتانسیل معدنی به دست آمده با استفاده از روش برش طبیعی که نقاط با تغییرات شدید را به عنوان حد تمایز انتخاب می کند، به چهار کلاس تقسیم بندی شدند. با بررسی مساحت هر کلاس نسبت به کل مساحت منطقه می توان راجع به نقشه های پتانسیل معدنی به دست آمده بهتر قضاوت کرد (شکل ۷). با توجه به شکل شماره هفت، می توان استنباط کرد که سیستم استنتاج فازی به دست آمده از الگوریتم ژنتیک درصد بالاتری از منطقه را به عنوان کلاس با پتانسیل خیلی کم معرفی می کند.



شکل ۷ مقایسه درصد مساحت کلاس های مختلف

در روش‌های به کار گرفته شده در این تحقیق وزن نقشه‌های معیار به طور مستقیم مشخص نیست؛ به این معنا که مدل‌سازی الگوهای داده به گونه‌ای پیچیده و غیرخطی مطرح می‌شود. بنابراین، در این قسمت تصمیم گرفته شد تا براساس میزان وابستگی نقشه‌های پتانسیل معدنی و نقشه‌های معیار، این میزان تأثیرات به نوعی برآورد شود. این میزان تأثیرات براساس ضریب همبستگی $\rho_{MPM,C}$ محاسبه شد (رابطه ۴).

$$\rho_{MPM,C} = \frac{\sigma_{MPM,C}}{\sqrt{\sigma_{MPM} \cdot \sigma_C}} \quad (4)$$

در رابطه بالا، $\sigma_{MPM,C}$ کوواریانس نقشه پتانسیل معدنی و نقشه معیار است. σ_{MPM} و σ_C به ترتیب انحراف معیار نقشه پتانسیل معدنی و نقشه معیار C است. جدول شماره سه نتایج به دست آمده را نشان می‌دهد.

جدول ۳ ضرایب همبستگی میان نقشه پتانسیل معدنی و نقشه‌های معیار

نقشه معیار	ضریب همبستگی نقشه پتانسیل معدنی	
	سیستم استنتاج فازی (خوشه‌بندی فازی)	سیستم استنتاج فازی (الگوریتم ژنتیک)
آلتراسیون	۰.۰۴۸	۰.۱۸۵
سنگ‌شناسی	۰.۳۶۵	۰.۲۸۰
فاصله تا گسل	۰.۲۹۰	۰.۰۹۹
شارژاییلیته	۰.۰۲۱	۰.۳۹۱
مقاومت ظاهری	۰.۲۹۱	۰.۰۳۰
فاکتور فلزی	۰.۴۰۲	۰.۰۶۳
شدت کل میدان مغناطیسی	۰.۴۶۱	۰.۵۰۱
مشق دوم شدت کل میدان مغناطیسی	۰.۱۹۱	۰.۰۰۹

نقشه‌های معیاری که ضریب همبستگی آن‌ها نزدیک به یک است، همبستگی بیشتری با نقشه پتانسیل معدنی دارد و به نوعی می‌توان وزن بیشتری را به آن‌ها اختصاص داد. همچنین،

نقشه‌های معیاری که ضریب هم‌بستگی آن‌ها نزدیک به صفر است، هم‌بستگی و وزن کمتری خواهند داشت. ضریب‌های هم‌بستگی مؤید این نکته‌اند که نقشه معیار شدت کل میدان مغناطیسی در هر دو روش بیشترین هم‌بستگی را با نقشه‌های پتانسیل معدنی به‌دست‌آمده دارد. به‌منظور ایجاد مبنایی قابل اعتماد برای ارزیابی نتایج روش‌های به‌کار گرفته‌شده، روشی ارائه شده که اصول آن بر محاسبه نسبت مقادیری که به‌وسیله نظریه آزمون آماری به‌دست می‌آیند، استوار است. برای این منظور چاه‌های اکتشافی موجود به دو دسته چاه‌های اکتشافی با پتانسیل بالا و پایین و نقشه‌های پتانسیل معدنی تهیه‌شده نیز به دو دسته پتانسیل معدنی کم و زیاد تقسیم شدند. برای هر چاه اکتشافی سه حالت مختلف ممکن تعریف شد:

۱- درست - مثبت (TP): تعداد چاه‌های اکتشافی با پتانسیل معدنی بالا/ پایین که به‌درستی در نواحی پیش‌بینی‌شده با پتانسیل معدنی زیاد/ کم قرار گرفته‌اند.

۲- غلط - مثبت (FP): تعداد چاه‌های اکتشافی با پتانسیل معدنی پایین که به‌غلط در نواحی پیش‌بینی‌شده با پتانسیل معدنی زیاد قرار گرفته‌اند.

۳- غلط - منفی (FN): تعداد چاه‌های اکتشافی با پتانسیل معدنی بالا که به‌غلط در نواحی پیش‌بینی‌شده با پتانسیل معدنی کم قرار گرفته‌اند.

پس از تعیین پارامترهای یادشده، نسبت‌های کامل بودن، درست بودن و کیفیت به‌منظور ارزیابی کمی نتایج به‌ترتیب با استفاده از رابطه‌های زیر محاسبه می‌شوند:

(۵)

$$Completeness = \frac{TP}{TP + FN}, \quad Correctness = \frac{TP}{TP + FP}, \quad Quality = \frac{TP}{TP + FP + FN}$$

جدول شماره چهار نتایج ارزیابی‌های کمی را که با روش یادشده محاسبه شده‌اند، نشان می‌دهد.

جدول ۴ نتایج ارزیابی‌های کمی

کیفیت	صحیح بودن	کامل بودن	پارامترهای ارزیابی کمی
۶۷.۴ درصد	۷۶.۳ درصد	۸۵.۳ درصد	سیستم استنتاج فازی (خوشه‌بندی فازی)
۶۷.۴ درصد	۱۰۰ درصد	۶۷.۴ درصد	سیستم استنتاج فازی (الگوریتم ژنتیک)

علی اصغر آل شیخ و همکاران استخراج پایگاه دانش فازی به منظور...

برای ارزیابی دقیق تر دقت در مدل سازی پتانسیل معدنی، از معیار مجموع جذر میانگین مربعات خطا برای داده های آموزشی ($RMSE_{Tr}$) و جذر میانگین مربعات خطا برای داده های کنترل ($RMSE_{Ch}$) استفاده شد. جدول شماره پنج مقادیر مجموع $RMSE_{Tr}$ و $RMSE_{Ch}$ را برای دو سیستم استنتاج فازی به دست آمده از خوشه بندی و الگوریتم ژنتیک نشان می دهد. همان طور که از جدول شماره پنج مشخص است، سیستم استنتاج فازی به دست آمده از الگوریتم ژنتیک و در نتیجه پایگاه دانش فازی موجود در آن دارای دقت بالاتری است.

جدول ۵ مقایسه کمی روش های مختلف براساس مجموع جذر میانگین مربعات خطا برای داده های آموزشی و داده های کنترل

روش	$RMSE_{Ch} + RMSE_{Tr}$
سیستم استنتاج فازی (خوشه بندی فازی)	۳۲.۹۷
سیستم استنتاج فازی (الگوریتم ژنتیک)	۱۶.۲۱

بررسی نتایج مطالعات و مقایسه آن ها با روش های مورد استفاده در این تحقیق نشان می دهد روش های استنتاج گری به دلیل دارا بودن ویژگی هایی مانند اعتماد پذیری، استحکام و صحت بالا دارای عملکرد بهتری هستند. فرایند پیشنهادی در این تحقیق مستقل از تعداد و نوع معیارهای ارائه شده است و می توان آن را با تغییرات لازم برای سایر مناطق و همچنین سایر کاربردها مانند رانش زمین به کار برد.

۶- تشکر و قدردانی

از همکاری صمیمانه شرکت ملی صنایع مس ایران به دلیل در اختیار قرار دادن داده های مورد نیاز سپاس گذاری می کنیم.

۷- منابع

- Alcalá-Fdez, J., R. Alcalá, M.J. Gacto, & F. Herrera, "Learning the Membership Function Contexts for Mining Fuzzy Association Rules by

Using Genetic Algorithms", *Fuzzy Sets and Systems*, 160(7), Pp. 905-921, 2009.

- Cordon, O., F. Herrera, F. Hoffmann & L. Magdalena, *Genetic Fuzzy Systems: Evolutionary Tuning and Learning of Fuzzy Knowledge Bases*, Singapore: World Scientific Pub Co Inc, 2001.
- Delaney, J., *Geographical Information Systems: An Introduction*, USA: Oxford University Press, 2000.
- Gen, M. & R. Cheng, *Genetic Algorithms and Engineering Optimization*, New York: Wiley- Interscienc, 2000.
- Herrera, F. & M. Lozano, "Fuzzy Evolutionary Algorithms and Genetic Fuzzy Systems: A Positive Collaboration between Evolutionary Algorithms and Fuzzy Systems", *Computational Intelligence*, Pp. 83-130, 2009.
- Herrera, F., M. Lozano & J.L. Verdegay, "A Learning Process for Fuzzy Control Rules Using Genetic Algorithms", *Fuzzy Sets and Systems*, 100(1-3), Pp. 143-158, 1998.
- Hosseinali, F. & A.A. Alesheikh, "Weighting Spatial Information in GIS for Copper Mining Exploration", *American Journal of Applied Sciences*, 5(9), Pp. 1187-1198, 2008.
- James, C., Bezdek, Robert Ehrlich, & W. Full, "FCM: The fuzzy C-Means Clustering Algorithms", *Computers & Geosciences*, 10(2-3), Pp. 191-203, 1984.
- Leite, E.P. & C.R.D.S. Filho, "Probabilistic Neural Networks Applied to Mineral Potential Mapping for Platinum Group Elements in the Serra Leste Region, Carajás Mineral Province, Brazil", *Computers & Geosciences*, 35(3), Pp. 675-687, 2008.



- Porwal, A.K., *Mineral Potential Mapping with Mathematical Geological Models*, Utrecht University, Enschede, Netherlands, 2006.
- Shad, R., A. Shad, M.S. Mesgari, H. Aghamohammadi & D. Molaei, "Fuzzy Topological Simulation for Deducing in GIS", *Applied Geomatics*, 1(4), Pp. 121-129, 2009.
- Siler, W. & J.J. Buckley, *Fuzzy Expert Systems and Fuzzy Reasoning*, New Jersey: John Wiley & Sons, Inc, 2005.
- Tahmasebi, P. & A. Hezarkhani, "Application of Adaptive Neuro-Fuzzy Inference System for Grade Estimation Case Study Sarcheshmeh Porphyry Copper Deposit, Kerman, Iran", *Australian Journal of Basic and Applied Sciences*, 4(3), Pp. 408-420, 2010.
- Tangestani, M.H. & F. Moore "The use of Dempster-Shafer Model and GIS in Integration of Geoscientific Data for Porphyry Copper Potential Mapping, North of Shahr-e-Babak, Iran", *International Journal of Applied Earth Observation and Geoinformation*, 4(1), Pp. 66-74, 2002.
- Vahidnia, M.H., A.A. Alesheikh, A. Alimohammadi & F. Hosseinali, "Landslide Hazard Zonation Using Quantitative Methods in GIS", *International Journal of Civil Engineering*, 7(3), Pp. 176-189, 2009.
- Wang, L.X. & J.M. Mendel, "Generating Fuzzy Rules by Learning from Examples", *IEEE Transactions on Systems, Man, and Cybernetics*, 22(6), Pp. 1414-1427, 1992.
- Wong, C.C. & S.M. Her, "A Self-generating Method for Fuzzy System Design", *Fuzzy Sets and Systems*, 103(1), Pp. 13-25, 1999.
- Yaochu, J., W. Von Seelen & B. Sendhoff, "On Generating FC3 Fuzzy Rule Systems from Data Using Evolution Strategies", *IEEE*

Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics, 29(6), Pp. 829-845, 1999.

- Zadeh, L.A. "Outline of a New Approach to the Analysis of Complex Systems and Decision Processes", *IEEE Transactions on Systems, Man, and Cybernetics*, 3(1), Pp. 28-44, 1973.



– Abstract

Weighting and integrating spatial information are vitally important in any analysis of the geospatial information systems (GIS). There are numerous sources of uncertainty in spatial phenomenon. Therefore, uncertainty in weighting spatial information must be considered. The purpose of this research is to construct a fuzzy knowledge base which contains all fuzzy rules along with membership functions to integrate uncertainty of spatial information in data integration. In the conventional design methods, the desired rules and functions are based on the expert's knowledge and experiences. However, we cannot perfectly represent the expert's knowledge by linguistic rules nor can we choose appropriate membership functions for fuzzy sets. Moreover, converting the experts' knowledge into if-then rules is difficult and often results in incompleteness, and conflicting knowledge. This is due to the fact that experts cannot express all their knowledge. These problems can be adjusted by applying techniques to construct fuzzy knowledge base from numerical input-output data. In this study, two methods of fuzzy clustering and genetic algorithm were employed to build fuzzy knowledge base automatically. For evaluating these two methods, the two extracted knowledge bases were used for mapping mineral potential mapping. The results indicated that knowledge base extracted by genetic algorithm with RMSE of 4 had better accuracy than fuzzy clustering.

Keywords: Fuzzy Inference System, Genetic Algorithm, Fuzzy Clustering, Mineral Potential Map, Geospatial Information System.