نشریه علمی پژوهشی مهندسی فناوری اطلاعات مکانی

سال سوم، شماره نخست، بهار ۱۳۹۴

Vol.3, No.1, spring 2015

۱ -۱٤

تعیین مناطق لرزه زمین ساخت با به‌کارگیری روش‌های خوشه‌بندی افرازی داده‌های لرزه‌ای

محمد کاظمی بیدختی[۱]، رحیم علی عباسپور[۲*]

۱-دانشجوی کارشناسی ارشد GIS - دانشکده مهندسی نقشه برداری و اطلاعات مکانی دانشگاه تهران

۲- استادیار دانشکده مهندسی نقشه برداری و اطلاعات مکانی دانشگاه تهران

تاریخ دریافت مقاله: ۱۳۹۴/۰۳/۱۶ تاریخ پذیرش مقاله: ۱۳۹۴/۰۸/۱۸

چکیده

مناطق لرزه زمین ساخت به مناطقی اطلاق می‌شوند که ویژگی‌های لرزه‌ای، زمین ساخت جنبا و سیستم گسلی شبیه به هم داشته باشند. شناسایی این مناطق به منظور ارزیابی خطر زمین لرزه، یکی از موضوعات حائز اهمیت در مطالعات لرزه‌ای است. روش‌های خوشه بندی به‌عنوان ابزاری مناسب برای شناسایی اطلاعات مرتبط باهم و استخراج الگوهای معنادار از بین تعداد زیادی از داده‌ها شناخته می‌شوند. با توجه به اینکه این روش‌ها علاوه بر نمایش بصری مناسب، نتایج قابل اعتمادی را نیز ارائه می‌دهند، می‌توانند ابزاری مناسب برای تعیین مناطق لرزه‌زمین ساخت باشند. در این تحقیق، از ۳ روش خوشه‌بندی افراز کا-مینز، ;کلارا و فازی سی-مینز استفاده شده است. به‌منظور تعیین تعداد بهینه خوشه‌ها در هر روش خوشه بندی، ۵ شاخص اعتبارسنجی دیویس بولدین، سیلوهت، اس دی، اس-دی بی دبلیو و شیه-بنی مورد استفاده قرار گرفت. پارامترهای موقعیت زمین لرزه، عمق زمین لرزه، بزرگا، جاذبه، مغناطیس و تراکم گسل‌ها به‌منظور تعیین تعداد مناطق لرزه‌زمین ساخت، که به‌عنوان ورودی روش‌های خوشه بندی مورد استفاده قرار می‌گیرند، در نظر گرفته شدند. پس از تعیین تعداد بهینه خوشه‌ها، برای ارزیابی نتایج روشهای خوشه بندی، از مدل لرزه زمین ساخت کمیته ملی سدهای بزرگ ایران استفاده شد. بر اساس نتایج بدست آمده مشخص گردید که روش کلارا در مقایسه با بقیه روش‌ها سازگاری بسیار بیشتری با مناطق لرزه‌زمین ساخت ایران را دارد. در نهایت پس از مقایسه روش‌های خوشه بندی در تعداد خوشه‌های بهینه با مدل لرزه-زمین ساخت مورد استفاده، تعداد مناطق لرزه‌زمین ساخت ۱۳ منطقه بدست آمد.

کلید واژه ها : مناطق لرزه‌زمین‌ساخت ، خوشه بندی، شاخص‌های اعتبارسنجی، داده‌های زلزله.

---

*نویسنده مسئول: خیابان کارگر شمالی (امیرآباد)، پردیس دانشکده‌های فنی دانشگاه تهران، دانشکده مهندسی نقشه‌برداری و اطلاعات مکانی.

تلفن: ۸۸۰۰۸۸۴۱

Email: abaspour@ut.ac.ir

۱- مقدمه

لرزه‌زمین‌ساخت [1] مطالعه ویژگی‌های لرزه‌خیزی، زمین ساخت جنبا و سیستم گسل‌ها در یک منطقه است[۱]. از این‌رو مناطقی را که چنین ویژگی‌های مشابه با هم داشته باشند، مناطق لرزه‌زمین‌ساخت می‌گویند[۲,۳]. پایه و مبنای ارزیابی خطر زمین‌لرزه[2]، شناسایی مناطق لرزه‌زمین‌ساخت می‌باشند[۴]. ارزیابی خطر زمین‌لرزه در زمینه‌های مختلفی کاربرد دارد. یکی از مرسوم‌ترین و پرکاربردترین آنها در طراحی سازه‌های مقاوم مسکونی در برابر زلزله می‌باشد. همچنین یکی از کاربردهای خاص آن در طراحی مقاوم زیرساخت‌های مهم از جمله نیروگاه هسته‌ای، پل، سد و پایگاه‌های نظامی می‌باشد[۵]. از این‌رو شناسایی مناطق لرزه‌زمین‌ساخت یکی از موضوعات مهم در مطالعات لرزه‌ای می‌باشد[۱].

در رابطه با تعیین تعداد مناطق لرزه‌زمین‌ساخت مطالعات مختلفی انجام شده است. در مطالعه [۶] مناطق لرزه‌زمین‌ساخت، برای شرق چین با استفاده از اطلاعات ژئودینامیکی، لرزه‌خیزی و زمین‌شناسی و الگوهای تنش بدست آمده است. کاظمی و جان [۷] پاکستان را به ۸ منطقه فعال تکتونیکی تقسیم‌بندی کرد. پس از آن، مونالیسا و همکاران [۸] پاکستان را به ۱۸ منطقه لرزه‌زمین‌ساخت با استفاده از لایه‌های اطلاعاتی مختلف تقسیم‌بندی کردند. کوماتو و ماساتاکا [۹] با استفاده از روش‌های آماری و روش PCA[3]، و با در نظرگرفتن پارامترهای جاذبه، لرزه زمین‌ساخت، گسل‌های فعال و لرزه‌خیزی تمامی کشور ژاپن را پهنه‌بندی کردند. در سال ۲۰۰۵، تیم تحقیقاتی USGS کشور افغانستان را به ۴ منطقه با در نظرگرفتن پارامترهای زمین‌شناسی، لرزه‌خیزی و گسل‌های فعال تقسیم‌بندی کردند[۱۰].

در ایران نیز به‌منظور تعیین مناطق لرزه‌زمین‌ساخت، مدل‌های مختلفی ارائه شده‌است. اولین مدل که توسط نوروزی در سال ۱۹۷۶ ارائه شد[3]، ایران با استفاده از اطلاعات جغرافیایی، ژئومورفولوژی و عوارض منطقه‌ای بر مبنای ۶۰۰ داده زلزله بین ۱۹۲۰ تا ۱۹۷۲ به ۲۳ منطقه تقسیم‌بندی شد. از مشکلات اصلی این مدل می‌توان به مطالعه بر روی تعداد بسیار کمی از داده‌های لرزه‌خیزی اشاره کرد. پس از آن مدل دیگری توسط بربریان در سال ۱۹۷۹ در نقد مدل نوروزی ارائه شد [۱۱] که در آن، ایران به ۹ منطقه بر مبنای اطلاعات زمین‌شناسی، لرزه زمین ساختی، ژئومورفولوژی و لرزه‌ای تقسیم بندی شده است. نوگل سعادت در سال ۱۹۹۳ [۱۲] مدل جدیدی برای مناطق لرزه‌زمین‌ساخت ایران پیشنهاد داد. در این مدل با بررسی پارامترهای مختلف تعداد این مناطق ۲۴ منطقه ارائه شد و پس از آن رمزی در سال ۱۹۹۵، مدلی را که تصحیح شده مدل نوگل سعادت بود، ارائه داد [۱۳]. در سال ۱۹۹۶ مدلی توسط کمیته ملی سدهای بزرگ ایران[۱۴][4] ارائه شد که در آن بر مبنای اطلاعات زمین‌شناسی، داده‌های به‌روز لرزه‌خیزی مانند مختصات زلزله‌ها، بزرگا و لایه‌های اطلاعاتی دیگر تعداد مناطق لرزه زمین ساختی برای ایران در نهایت ۱۶ منطقه بدست آمد. از آخرین مطالعات دهه ۹۰ میلادی، می‌توان به میرزایی و همکاران [۱۵] و توکلی و آشتیانی [۱۶] اشاره کرد. از سال ۲۰۰۰ به بعد تقریباً مطالعات در حوزه لرزه‌زمین‌ساخت انجام نشد تا سال ۲۰۰۸ که انصاری و همکاران [۱] تحلیل خوشه‌ای بر روی داده‌های زلزله انجام دادند. آنها در این مقاله نشان دادند تحلیل‌های حاصل از خوشه‌بندی سازگاری بالایی با مناطق لرزه‌زمین‌ساخت دارند. در این تحقیق تنها پارامتر موقعیت زمین‌لرزه‌ها

1 Seismotectonic
2 Seismic Hazard Assessment
3 PCA : Principle Componenet Analysis
4 Iranian National Committee on Large Dams(IRCOLD)

در نظر گرفته شده است و فرض شده است که بین سایر پارامترها همبستگی مکانی وجود دارد. علاوه بر این، زمانی و همکاران در سال ۲۰۰۹ [۱۷] و مجرب و همکاران در سال ۲۰۱۳ [۱۸] مطالعاتی در تعیین مناطق لرزه زمین ساخت با استفاده از روش SOM[1] انجام دادند.

با توجه به اینکه برای ارائه هر مدل لرزه‌زمین‌ساختی از داده‌های مختلف با سطح دقت متفاوت و همچنین از پارامترهای مختلف استفاده شده است، لذا مدل‌های گوناگونی براساس نظر کارشناسان به منظور تعیین تعداد مناطق لرزه‌زمین‌ساخت ارائه شده‌اند. از این‌رو در حال حاضر مدل جامعی برای مناطق لرزه زمین‌ساخت وجود ندارد. با توجه به پیشرفت تکنولوژی و در راستای آن بالاتر رفتن سطح دقت دستگاه‌های اندازه‌گیری داده‌های لرزه‌ای، استفاده از داده‌های به‌روز در کنار داده‌های تاریخی، همچنین در نظر گرفتن پارامترهای مناسب و استفاده‌از روش‌های کارآمد، گامی اساسی در راستای ارائه مدل جامع مناطق لرزه‌زمین‌ساخت خواهد بود.

داده‌کاوی عبارت است از استخراج اطلاعات و دانش و کشف الگوهای پنهان از پایگاه داده‌های بسیار بزرگ[۱۹]. با توجه به پیشرفت دستگاه‌های اندازه‌گیری داده‌های زلزله و تولید تعداد زیاد داده بوسیله آنها، لذا استفاده از روش‌های قدرتمند که علاوه بر نمایش بصری مناسب، توانایی ارتباط با تعداد زیادی داده را دارند حائز اهمیت می‌باشد[۱]. لذا با توجه به حجم زیاد داده‌های لرزه‌ای، روش‌های خوشه بندی که علاوه بر نمایش بصری مناسب، توانایی استخراج الگوهای پنهان معنادار از تعداد زیاد داده را نیز دارند، ابزاری کارا در تعیین مناطق لرزه‌زمین‌ساخت خواهند بود.

خوشه بندی به‌عنوان یکی از روش‌های داده‌کاوی و کشف دانش، تکنیکی برای گروه بندی مشاهدات به تعداد مشخصی از خوشه‌های مختلف است، به طوریکه مشاهدات واقع در یک خوشه بیشترین شباهت را باهم و مشاهدات خوشه‌های مختلف بیشترین تفاوت ممکن را با یکدیگر داشته‌باشند[۲۰] خوشه‌بندی کاربردهای فراوانی در علوم مختلف از جمله بیولوژی [۲۱]، امنیت [۲۲]، تشخیص الگوی تصاویر [۲۳]، هوش تجاری [۲۴] و جستجوی تحت وب [۲۵] دارد.

بر مبنای اینکه از چه روش خوشه‌بندی استفاده شود، خوشه‌بندی‌های مختلفی بر روی مجموعه یکسانی از داده‌ها خواهیم داشت [۲۶]. تاکنون، الگوریتم‌های خوشه‌بندی متعددی توسط پژوهشگران در حوزه‌های مختلف ارائه شده‌اند؛ هرچند به علت همپوشانی نمی‌توان تقسیم بندی جامع و دقیقی از کلیه روش‌های خوشه بندی موجود ارائه داد، با اینحال می‌توان روش‌های خوشه‌بندی را به دو دسته افرازی و سلسله مراتبی تقسیم بندی نمود[۲۷,۲۸].

روش‌های افرازی، مجموعه‌ای با n عضو را به k گروه افراز می‌کند بطوریکه $k \leq n$ و هر گروه مشخص کننده یک خوشه می‌باشد[۲۹]. روشهای سلسله مراتبی مجموعه داده‌ها را به دسته‌ها و زیر دسته‌هایی بصورت درختی تقسیم می‌کند. این روش‌ها به‌منظور حل برخی از معایب روش‌های افرازی از جمله غیرقطعی بودن آن ارائه شدند[۳۰]. معایب اصلی این روش‌ها عبارتند از: ۱) انتخاب معیارهای درست برای شرط توقف سخت است. ۲) بیشتر الگوریتم‌های سلسله مراتبی، خوشه‌ها را اصلاح نمی‌کنند و مجددا این خوشه‌ها تشکیل می‌شوند. از طرفی روش‌های افرازی، با توجه به اینکه روندی تکراری دارند لذا به تدریج در آنها خوشه‌ها بهبود می‌یابند. لذا با داشتن داده‌های مناسب، نتایج خوشه‌بندی‌ها از کیفیت بالایی برخوردار خواهند بود. همچنین در صورت زیاد بودن

1 SOM : Self-Organizing Map

تعداد متغیرها، این روش نسبت به روش سلسله مراتبی دارای سرعت بالاتری می‌باشد. [۲۷].

در این تحقیق، از ۳ روش خوشه‌بندی افرازی کا-مینز، کلارا[1] و فازی سی-مینز استفاده شده است. روش کا-مینز، به دلیل مزایایی مثل سهولت پیاده‌سازی، سادگی و کارایی بالا، به طور وسیعی مورد استفاده قرار می‌گیرد[۳۱]. روش کلارا در مقایسه با روش کا-مینز علاوه بر اینکه نسبت به نویز از حساسیت کمتری برخوردار است همچنین حساسیت آن به داده‌های دور افتاده[2] نیز بسیار کمتر است[۳۲]. در روش فازی سی-مینز خوشه‌بندی به صورت فازی انجام می‌شود. در این مقاله باتوجه به اینکه مرزمناطق لرزه‌زمین‌ساخت به صورت دقیق مشخص نیست، لذا یک روش خوشه بندی فازی همانند فازی سی-مینز، که رایج‌ترین روش خوشه‌بندی فازی محسوب می‌شود[۳۳]، نیز انتخاب شده‌است. با توجه به اینکه در این تحقیق از تعداد بسیار زیادی داده استفاده شده‌است لذا روش‌های مورد استفاده بایستی برای داده‌ها با حجم بسیار زیاد نیز مناسب باشند که هر ۳ روش مورد استفاده در این مقاله نیز این ویژگی را دارند[۳۱].

یکی از چالش‌های اساسی در روش‌های خوشه بندی افرازی، تعیین تعداد بهینه خوشه‌ها می‌باشد[۳۲]. بدین منظور برای حل این مسئله از چندین شاخص اعتبارسنجی[3] استفاده شده است. استفاده همزمان از چندین شاخص اعتبارسنجی و مقایسه نتایج آنها با همدیگر در تعیین تعداد بهینه خوشه‌ها توصیه شده‌است. [۳۴,۳۵]

در ادامه، در بخش ۲ مروری بر خوشه‌بندی داده‌ها و ۵ روش خوشه‌بندی مورد استفاده در این تحقیق انجام شده‌است. در بخش ۳، شاخص‌های اعتبارسنجی به‌صورت خلاصه معرفی شده اند. در بخش ۴ مدل لرزه‌زمین‌ساخت مورد استفاده و در بخش ۵ منطقه مورد مطالعه تشریح شده‌است. پیاده‌سازی و ارزیابی نتایج در بخش ۶ و در نهایت نتیجه‌گیری در بخش ۷ صورت گرفته‌است.

۲- مروری بر خوشه‌بندی داده‌ها

خوشه‌بندی، یکی از شاخه‌های یادگیری بدون نظارت بوده و به فرآیند قرار گرفتن داده‌ها در چندین گروه گفته می‌شود، بطوریکه داده‌های داخل یک گروه، بیشترین شباهت را با هم داشته و با داده‌های گروه‌های دیگر کمترین شباهت را داشته باشند.

برای شباهت می توان معیارهای مختلفی را در نظر گرفت. به عنوان مثال، یکی از معیارها فاصله می‌باشد که بر اساس آن نقاطی که به یکدیگر نزدیکتر هستند، به‌عنوان یک خوشه در نظر گرفته می‌شوند (خوشه‌بندی مبتنی بر فاصله). استفاده از روش‌های خوشه‌بندی متفاوت باعث ایجاد خوشه‌های متفاوت بر روی یک مجموعه داده یکسان خواهد شد[۲۶].

در ادامه ۳ روش خوشه‌بندی افرازی مورد استفاده در این مقاله شامل کا-مینز، کلارا و فازی سی-مینز معرفی شده‌اند.

۲-۱- روش کا-مینز

این روش برخلاف سادگی در محاسبات، بیشترین استفاده را در روش‌های خوشه‌بندی افرازی دارد [۳۲]. در این روش، خوشه‌بندی بر مبنای کمینه‌کردن تابع هدف فاصله اقلیدسی صورت می‌گیرد[۳۶]. این روش که روالی تکراری دارد، برای تعدادی ثابت از خوشه‌ها سعی در تخمین موارد زیر را دارد:

۱. بدست آوردن نقاطی به عنوان مراکز خوشه‌ها که این نقاط در واقع همان میانگین نقاط متعلق به هر خوشه هستند.

1 Clustering Large Applications
2 Outlier
3 Validity Indices

۲. نسبت دادن هر نمونه داده به یک خوشه که کمترین فاصله تا مرکز آن خوشه را دارا باشد.

تابع هدف فاصله به صورت زیر تعریف می‌شود:

$$J = \sum_{i=1}^{c}\sum_{j=1}^{n} ||X_j^{(i)} - C_i||^2$$

رابطه (۱)

که در این رابطه n تعداد نقاط، c تعداد خوشه‌ها، C_i مرکز خوشه iام و $X_j^{(i)}$ نقطه jام در خوشه iام می‌باشد.

از معایب این الگوریتم می‌توان به موارد زیر اشاره کرد[۳۱]:

۱) حساس بودن نسبت به داده‌های دور از مرکز
۲) وابسته بودن نتایج به انتخاب مراکز اولیه
۳) ناتوانی در شناسایی داده نویز

۲-۲- روش کلارا

این الگوریتم که توسط [۳۷] در سال ۱۹۹۰ ارائه شد، توسعه یافته روش K-medoid می‌باشد. روش K-medoid به تعداد تکرار زیادی نیاز دارد و به همین دلیل برای خوشه‌بندی مجموعه داده‌های حجیم مناسب نیست. از این‌رو روش کلارا معرفی شده است. این روش که برای مجموعه داده‌های بزرگ پیشنهاد شده است چندین نمونه تصادفی از پایگاه داده برمی‌دارد و الگوریتم PAM[1] را روی هر نمونه اجرا کرده و آن نمونه را خوشه‌بندی می‌کند. بهترین خوشه‌بندی به‌عنوان خروجی ارائه می‌شود. سپس عناصر باقیمانده پایگاه داده را بر مبنای این خروجی به نزدیکترین خوشه اختصاص می‌دهد. در واقع در این الگوریتم یک قسمت کوچکی از داده‌های واقعی به‌عنوان نماینده داده‌ها انتخاب می‌شوند[۳۷].

۲-۳- روش فازی فازی سی-مینز

الگوریتم‌های خوشه‌بندی قطعی[2]، داده‌ها را به‌گونه‌ای افراز می‌کنند که هر داده دقیقاً به یک خوشه تخصیص داده شود. دربرخی موارد نمی‌توان هر داده را دقیقاً به یک خوشه تخصیص داد چراکه برخی داده‌ها بین خوشه‌ها قرار می‌گیرند. در این موارد روش‌های خوشه بندی فازی ابزارهایی مناسبتر برای نمایش ساختار واقعی این نوع داده‌ها هستند[۳۲]. در این تکنیک هر شی با درجه عضویتی به خوشه‌ها تعلق می‌گیرد.

از مزایای این روش می‌توان به قابلیت همیشه همگرا شدن و همچنین بدون نظارت بودن آن اشاره کرد. از طرفی حساسیت به حدس‌های اولیه و حساسیت به نویز از معایب این روش محسوب می‌شود.

این روش که توسعه یافته روش کا-مینز می‌باشد تابع هدف در آن به صورت زیر تعریف می‌شود:

$$J = \sum_{i=1}^{c}\sum_{j=1}^{n} u_{ij}^{m} d_{ij}^{2} = \sum_{i=1}^{c}\sum_{j=1}^{n} u_{ij}^{m} \| x_j - c_i \|$$

رابطه (۲)

در این رابطه، u_{ij} درجه عضویت نقطه iام در خوشه jام، $m \in (1,+\infty)$ که به آن توان وزنی گفته می‌شود و d_{ij} فاصله اقلیدسی نقطه iام تا مرکز خوشه jام می‌باشد.

همانند روش کا-مینز، در این روش نیز به دنبال کمینه کردن رابطه(۲) هستیم که در آن درجه عضویت نقطه iام در خوشه jام به صورت زیر بدست می‌آید:

$$u_{ij} = \frac{1}{\sum_{k=1}^{c} \left(\frac{d_{ij}}{d_{kj}}\right)^{2/(m-1)}}$$

رابطه (۳)

1 Partitioning Around Medoids

2 Hard Clustering

۳- مروری بر شاخص‌های اعتبارسنجی

یکی از موضوعات اساسی در خوشه‌بندی، ارزیابی نتایج حاصل از الگوریتم‌های خوشه‌بندی است[۳۸]. بدین منظور شاخص‌هایی به منظور ارزیابی خوشه‌بندی ارائه شده‌اند. شاخص‌های اعتبارسنجی بر مبنای بیشترین شباهت بین اعضای هر خوشه و بیشترین تفاوت هر خوشه با خوشه‌های دیگر، محاسبه می‌شوند. در همین راستا شاخص‌های زیادی ارائه شده است. در [۳۲] مقایسه‌ای بین تعداد بسیار زیادی شاخص بر روی یک مجموعه داده آزمایشی انجام گرفت و در نهایت نتایج آنها در جدولی ارائه شد. با توجه به نتایج حاصل از این تحقیق، ۵ شاخص

دیویس-بولدین[1]، سیلوهت[2]، اس دی[3]، اس-دی بی دبلیو[4] و شیه-بنی[5] در مقایسه با سایر شاخص‌ها نتایج بهتری ارائه داده بودند. لذا برای تعیین تعداد بهینه خوشه‌ها در این مقاله از این ۵ شاخص استفاده شده است. در جدول ۱ توضیح مختصری از ۵ شاخص مورد استفاده به همراه حالت بهینه تعداد خوشه‌ها در هر شاخص، آمده است.

تمامی الگوریتم‌های خوشه‌بندی و شاخص‌های اعتبارسنجی مورد استفاده در این مقاله در نرم‌افزار R پیاده‌سازی شده‌اند.

۴-مدل لرزه زمین ساخت مورد استفاده

اگرچه مدل‌های فراوانی به‌منظور تعیین مناطق لرزه زمین ساخت ارائه شده است ولی تا به حال مدل جامعی بدین منظور ارائه نشده است. مهمترین علل این امر تفاوت در تعداد و دقت داده‌های مورد استفاده و همچنین تعداد پارامترهای موثر می‌باشد.

جدول ۱- توضیح مختصری از شاخص‌های اعتبارسنجی مورد استفاده

شاخص	توضیح	حالت بهینه
دیویس بولدین	این شاخص از شباهت بین دو خوشه استفاده می‌کند که بر اساس پراکندگی یک خوشه و عدم شباهت بین دو خوشه تعریف می‌شود. این شاخص در واقع میانگین شباهت بین هر خوشه با شبیه‌ترین خوشه به آن را محاسبه می‌کند [۳۹].	کمترین
سیلوهت	این شاخص بر مبنای اندازه گیری تراکم و جدایی خوشه‌ها می‌باشد بطوریکه میزان شباهت داده‌های یک خوشه با مقایسه داده‌های خوشه‌های دیگر محاسبه می‌شود [۴۰].	بیشترین
اس‌دی	اساس این شاخص، پراکندگی و جدایی کلی خوشه‌ها است. پراکندگی از طریق محاسبهٔ واریانس خوشه‌ها و واریانس کل محموعهٔ داده‌ها بدست می‌آید.[۴۱]	کمترین
اس-دی‌بی دبلیو	همانند شاخص اس‌دی این معیار هم بر اساس تراکم درون‌خوشه‌ای و میزان جدایی خوشه‌ها می‌باشد اما در این شاخص سعی شده تا چگالی خوشه‌ها نیز دخیل شود[۴۲].	کمترین
شیه-بنی	میزان تراکم و جدایی خوشه‌ها و همچنین مجموع اختلافات بین خوشه‌ها و جدایی آنها را نیز تعیین می‌کند [۴۳]	کمترین

در این تحقیق با توجه به اینکه هدف تعیین مناطق لرزه‌زمین‌ساخت با استفاده از روش‌های خوشه‌بندی افرازی می‌باشد، از این‌رو به منظور ارزیابی نتایج بدست آمده، بایستی خروجی های حاصل با مناسبترین مدل لرزه‌زمین‌ساخت موجود مقایسه شود. مدل

[1] Davies-Bouldin
[2] Silhouette
[3] SD
[4] S_Dbw
[5] Xie-Beni

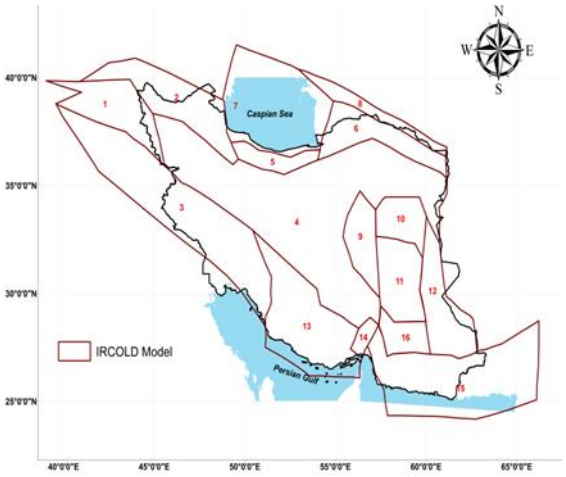
لرزه زمین ساختی که در این مقاله استفاده شده است، مدل کمیته ملی سدهای بزرگ ایران می‌باشد (شکل ۱). این مدل علاوه بر اینکه جزء جدیدترین مدل‌های رایج و معتبر لرزه‌زمین‌ساختی شناخته می‌شود، در آن از داده‌های به‌روز با دقت مناسب به همراه داده‌های تاریخی استفاده شده‌است. همچنین اکثر پارامترهای موثر در تعیین مناطق لرزه زمین‌ساختی در این مدل نیز در نظر گرفته شده است[۱۴]. باتوجه به شکل ۱، تعداد مناطق لرزه‌زمین‌ساخت ایران در این مدل ۱۶ منطقه می‌باشد که نام هر منطقه در جدول ۲ آمده است.

جدول ۲- اسامی مناطق لرزه‌زمین ساخت مدل کمیته ملی سدهای بزرگ ایران

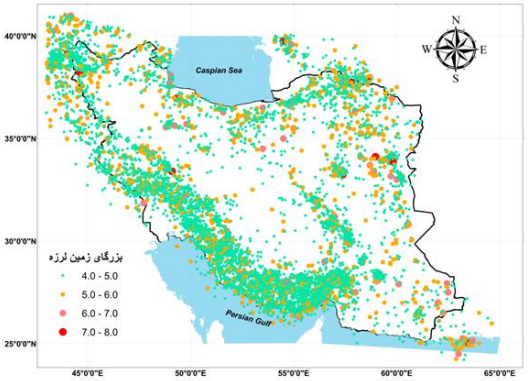
نام منطقه	شماره منطقه	نام منطقه	شماره منطقه
آذربایجان غربی	۱	طبس	۹
آذربایجان شرقی	۲	فردوس	۱۰
شمال غربی زاگرس	۳	لوت	۱۱
مرکز ایران	۴	نهبندان	۱۲
البرز	۵	جنوب شرقی زاگرس	۱۳
کوپه داغ	۶	خورگو	۱۴
جنوب دریای خزر	۷	مارکان	۱۵
شمال کوپه داغ	۸	جازموریان	۱۶

۵- داده‌های مورد مطالعه

در این مقاله، از کاتالوگ لرزه‌ای جامع و همگنی که توسط شاهور و همکاران [۴۴] برای فلات ایران در بازه زمانی ۱۹۰۰-۲۰۱۱ فراهم شده، استفاده شده است. در این تحقیق به منظور تعیین مناطق لرزه‌زمین‌ساخت به طورکلی از ترکیبی از اطلاعات لرزه‌ای، توپوگرافی، ژئوفیزیک و زمین‌شناسی با توجه به مطالعات پیشین انجام شده در این حوزه استفاده شده‌است. در این راستا پارامترهایی که به‌عنوان ورودی روش‌های خوشه‌بندی در نظر گرفته شده‌اند عبارتند از: موقعیت زمین‌لرزه، عمق زمین‌لرزه، بزرگا، گرانش، خاصیت مغناطیسی زمین و تراکم گسل‌ها. تراکم گسل‌ها از روش چگالی کرنل بدست آمد و سپس برای هر زمین‌لرزه مقدار آن درونیابی شد. شکل ۲ توزیع مختصاتی داده‌های مورد استفاده را به همراه بزرگی آن‌ها نشان داده‌است. شاهور و همکاران به‌منظور تست کامل بودن داده‌های لرزه‌ای، زلزله‌ها با بزرگای بیشتر از ۴ ریشتر را در نظر گرفتند.



شکل ۱- مدل لرزه زمین ساختی کمیته ملی سدهای بزرگ ایران



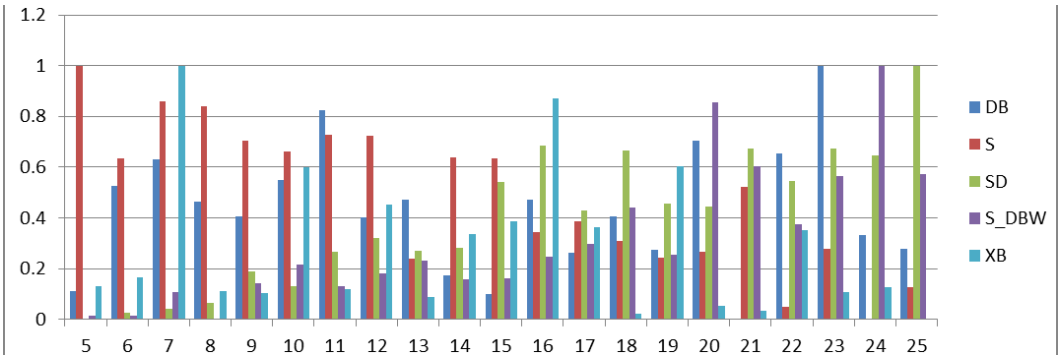
شکل ۲- توزیع مختصاتی داده‌های لرزه‌ای ایران از سال ۲۰۱۲ – ۱۹۰۰ به همراه بزرگای آنها بر حسب ریشتر

۶- پیاده‌سازی و ارزیابی نتایج

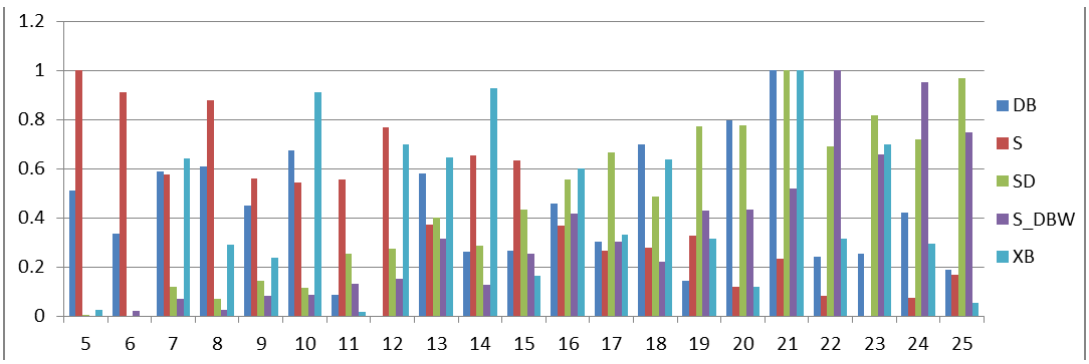
در ابتدا بایستی تعداد بهینه خوشه‌ها در هر ۳ روش خوشه‌بندی تعیین شوند. بدین منظور از شاخص‌های اعتبارسنجی که در بخش ۳ اشاره شد استفاده می‌شود.

لازم به ذکر است که هیچ شاخصی به تنهایی نمی‌تواند برای ارزیابی خوشه‌بندی‌ها مورد استفاده قرار گیرد. بدین منظور در این تحقیق از ۵ شاخص استفاده شده است که با مقایسه نتایج آنها با هم، تعداد خوشه‌های بهینه، که مشخص کننده‌ی مناطق لرزه‌زمین‌ساخت می‌باشند، بدست می‌آیند. با توجه به اینکه کمترین تعداد مناطق لرزه‌زمین‌ساخت ارائه شده ۵ منطقه [۱۵] و بیشترین آن۲۳ منطقه [۳] می‌باشد، از این‌رو تعداد خوشه‌ها بین ۵ تا ۲۵ در نظر گرفته شده‌است. شکل‌های ۳ تا ۵ مقادیر شاخص‌های اعتبارسنجی برای هر روش خوشه‌بندی از ۵ تا ۲۵ خوشه را نشان می‌دهند. با مقایسه مقادیر شاخص‌های اعتبارسنجی بدست آمده برای هر روش خوشه‌بندی، تعداد خوشه‌های بهینه برای هر روش به‌دست می‌آیند که در جدول ۳ مشخص شده‌است.

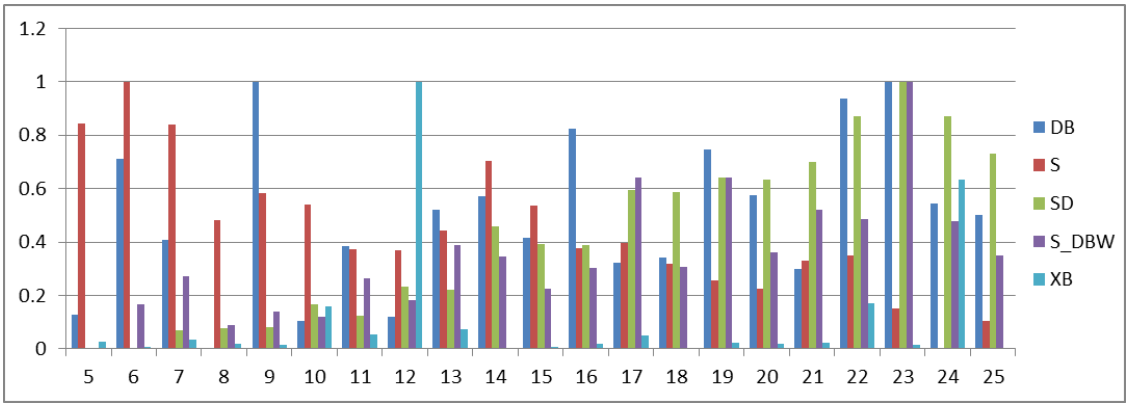
سپس به منظور ارزیابی نتایج، خوشه‌بندی باتوجه به تعداد خوشه‌های بهینه بدست آمده، انجام می‌شود و سپس نتایج هر روش خوشه‌بندی با مدل لرزه زمین ساخت که در بخش ۲ معرفی شد، مقایسه می‌شوند. شکل‌های ۶ (الف ، ب و ج) ، ۷ (الف، ب و ج) و ۸ (الف، ب و ج) نتایج حاصل از خوشه‌بندی روش‌های کا-مینز، کلارا و فازی سی-مینز براساس تعداد خوشه‌های بهینه بدست آمده به همراه مدل لرزه-زمین‌ساخت مورد استفاده را نشان می‌دهند. با توجه به این اشکال، هر منطقه لرزه‌زمین‌ساخت به صورت مجزا با خروجی حاصل از روش‌های خوشه‌بندی در تعداد بهینه خوشه‌ها، مقایسه می‌گردد تا مشخص شود که کدام روش سازگاری بهتری با آن منطقه دارد. به عنوان نمونه در روش کا-مینز در $K_{opt}=8$، تنها منطقه پنجم بوسیله این روش شناسایی شده است و سایر مناطق با دیگر خوشه‌ها هیچگونه سازگاری ندارند اما در همین روش در $K_{opt}=13$، علاوه بر منطقه پنجم، مناطق ۶، ۹، ۱۰، ۱۱ و ۱۵ نیز توانسته از خوشه‌های بدست آمده، استخراج گردند. به همین ترتیب این کار برای دیگر روش‌ها در تمامی K_{opt}ها انجام می‌شود که در نهایت نتایج در جدول ۴ آمده است.



شکل ۳ – مقادیر شاخص های اعتبارسنجی روش کا-مینز از ۵ تا ۲۵ خوشه



شکل ۴- مقادیر شاخص های اعتبارسنجی روش کلارا از ۵ تا ۲۵ خوشه

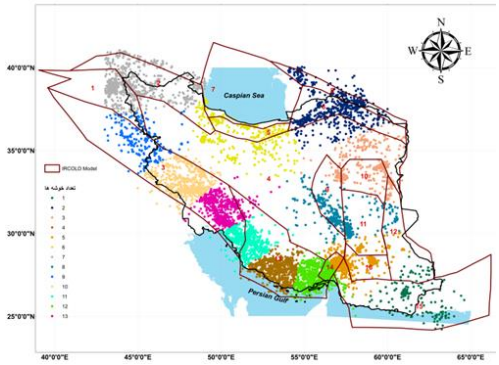


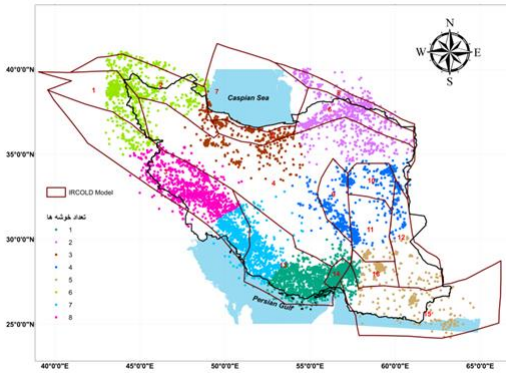
شکل ۵- مقادیر شاخص های اعتبارسنجی روش فازی سی-مینز از ۵ تا ۲۵ خوشه

جدول ۳ – تعداد خوشه‌های بهینه در هر روش خوشه‌بندی

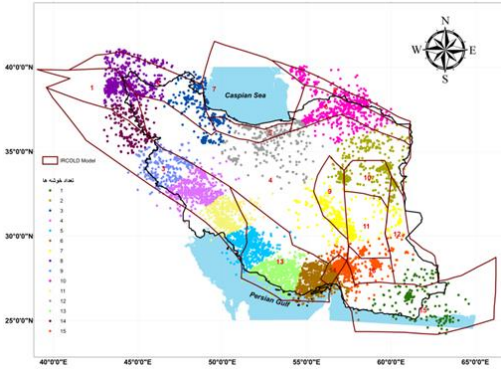
روش خوشه‌بندی	تعداد بهینه خوشه‌ها (K_{opt})
کا-مینز	۸, ۱۳, ۱۵
کلارا	۸, ۱۲, ۱۳
فازی سی-مینز	۶, ۸, ۱۴

(ب)





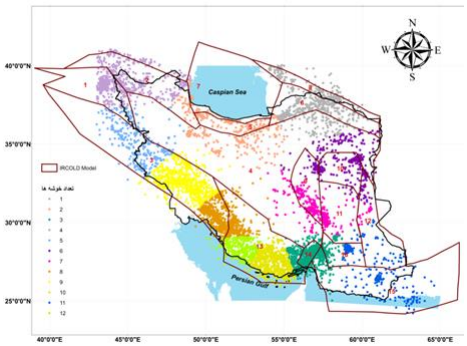
(الف)



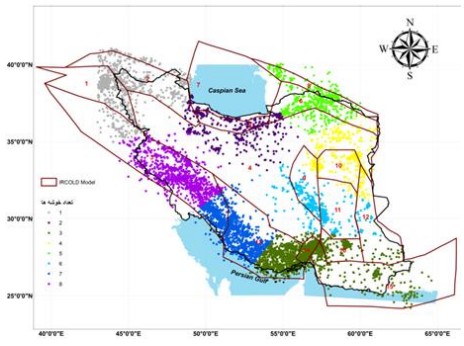
(ج)

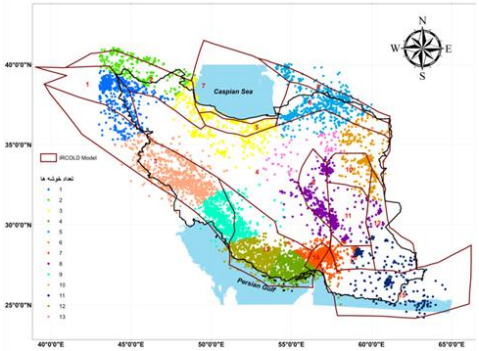
شکل ۶ الف- خوشه‌بندی به روش کا-مینزدر Kopt=۸ , ب- خوشه‌بندی به روش کا-مینزدر Kopt=۱۳ ,
ج-خوشه بندی به روش کا-مینزدر Kopt=۱۵

(ب)



(الف)

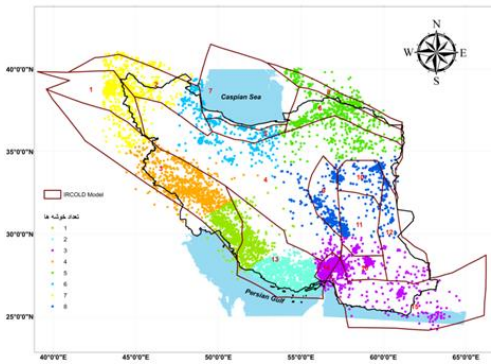




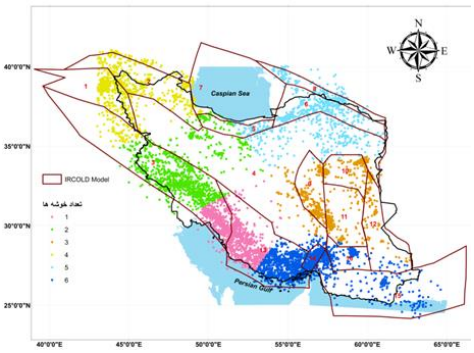
(ج)

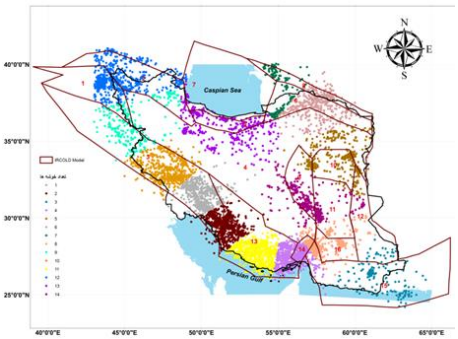
شکل ۷ الف- خوشه‌بندی به روش کلارا در Kopt=۸ , ب- خوشه‌بندی به روش کلارا در Kopt=۱۲ ,
ج- خوشه بندی به روش کلارا در Kopt=۱۳

(ب)



(الف)





(ج)

شکل ۸ الف- خوشه‌بندی به روش فازی سی-مینزدر Kopt=۶ , ب- خوشه‌بندی به روش فازی سی-مینزدر Kopt=۸ ,
ج- خوشه‌بندی به روش فازی سی-مینزدر Kopt=۱۴

جدول ۴- مقایسه روش‌های خوشه بندی کا-مینز، کلارا و فازی سی-مینز در تعداد بهینه خوشه‌ها (Kopt) با ۱۶ منطقه لرزه زمین‌ساخت مدل کمیته ملی سدهای بزرگ ایران

	K_{opt}	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰	۱۱	۱۲	۱۳	۱۴	۱۵	۱۶
کا-مینز	۸					✓											
	۱۳					✓	✓		✓	✓	✓	✓				✓	
	۱۵						✓		✓	✓	✓	✓				✓	
کلارا	۸					✓	✓		✓	✓	✓	✓					
	۱۲		✓			✓	✓		✓	✓	✓	✓			✓	✓	
	۱۳	✓	✓			✓	✓		✓	✓	✓	✓			✓	✓	
فازی سی-مینز	۶																
	۸					✓	✓		✓								
	۱۴					✓				✓	✓	✓				✓	

بر اساس جدول ۴ : ۱) روش کلارا در مقایسه با دو روش دیگر سازگاری بسیار بیشتری با مدل کمیته ملی سدهای بزرگ ایران دارد. ۲) بیشترین مناطق لرزه زمین ساخت شناسایی شده توسط روش کلارا در Kopt =۱۳ بوده‌است. ۳) مناطق ۴، ۷ و ۱۲و ۱۶ به دلیل فقدان داده‌های لرزه‌ای کافی، توسط هیچکدام از روش‌های خوشه‌بندی شناسایی نشده‌اند. ۴) در روش کلارا در Kopt =۱۳ در منطقه زاگرس، باوجود اینکه مناطق ۳ و ۱۳ توسط این روش شناسایی نشده‌اند، ولی ۴ خوشه بدست آمده برای این منطقه در این روش، سازگاری مناسبی با منطقه زاگرس مدل نوروزی [۳] دارند.

۷- نتیجه گیری

در این مقاله، به منظور تعیین تعداد مناطق لرزه زمین ساخت فلات ایران از روش‌های خوشه‌بندی افرازی شامل: کا-مینز ، کلارا و فازی سی مینز استفاده شد. بدین‌منظور ابتدا با استفاده از شاخص‌های اعتبارسنجی تعداد بهینه خوشه‌ها برای هر روش به‌دست آمد. سپس به‌منظور ارزیابی نتایج روش‌های خوشه بندی در تعداد بهینه خوشه‌ها، از مدل لرزه زمین ساختی کمیته ملی سدهای بزرگ ایران استفاده شد. در نهایت پس از مقایسه مناطق مدل لرزه‌زمین‌ساخت مورد استفاده با خروجی حاصل از روش‌های خوشه‌بندی در تعداد بهینه خوشه‌ها، مشخص شد که به‌طور کلی روش کلارا در مقایسه با دو روش دیگر سازگاری بسیار بیشتری با مدل لرزه زمین ساخت دارد. مناطق مرکزی ایران، جنوب دریای خزر، نهبندان و جازموریان، به دلیل نبود داده‌های کافی، هیچکدام از روشها قادر به شناسایی این مناطق نبودند. در نهایت در روش خوشه‌بندی کلارا در حالتی که K_{opt}= ۱۳ می‌باشد بیشترین تعداد مناطق مدل لرزه‌زمین‌ساخت کمیته ملی سدهای بزرگ ایران شناسایی شده است. به‌عبارتی این روش در این تعداد خوشه‌های بهینه در مقایسه با تمامی حالت‌های دیگر بیشترین سازگاری با مناطق لرزه زمین ساخت ایران دارد. از این‌رو، تعداد مناطق لرزه‌زمین‌ساخت برای ایران ۱۳ منطقه پیشنهاد می‌شود.

مراجع

[1] A. Ansari, A. Noorzad, H. Zafarani,"Clustering analysis of the seismic catalog of Iran", Computers & Geosciences, Vol. 35, Issue 3, PP. 475 – 486,2009.

[2] M.Berberian. "Contribution to Seismotectonics of Iran (part II)" Geological Survey of Iran, Report no. 39, PP.518, 1976.

[3] A. Nowroozi. "Seismotectonic provinces of Iran", Bulletin of the Seismological Society of America, Vol. 66, No. 4, PP. 1249-1276,1976.

[4] C.Meletti, E. Patacca, P. Scandone "Construction of a Seismotectonic model: the case of Italy". Pure Appl Geophys, Vol.157 , Issue 1-2,:pp. 11-35, 2000.

[5] J.Kulgel. " Seismic hazard analysis —Quo vadis?" Earth Science Reviews, Vol.88, Issues 1–2, PP. 1–32,2008.

[6] H.Ye, G.Chen,Q. Zhou "Study on the intraplate potential seismic sources". In: Proceedings of the Fifth International Conference on Seismic Zonation, Nice, France, pp. 1424–1430, 1995.

[7] A.H.Kazmi, M.Q.Jan. "Geology and Tectonics of Pakistan". Graphic Publishers, Karachi, vol.64,pp. 554,1997.

[8] M.Q.MonaLisa,Jan, A.A.Khwaja "A preliminary seismotectonic zonation map of the Nw himalayan fold-and-thrust belt, Pakistan, for the period 1904–2006 ". In: Proceedings of the Pakistan Academy of Science, vol. 46, pp. 175–182,2009.

[9] T.Kumamoto, T.Masataka, "Multivariate statistical analysis for seismotectonic zonation by the use of earthquake, active fault and crustal structure". Japan Geosciences Union Meeting, 2011.

[10] R.L.Wheeler, C.G.Bufe, M.L.Johnson, R.L.Dart. "Seismotectonic Map of Afghanistan, with Annotated Bibliography". U.S. Geological Survey, Open-File Report. p 31, 2005.

[11] M.Berberian. "Discussion of the paper A. A. Nowroozi, 1976 "Seismotectonic provinces of Iran" ".Bulletin of the Seismological Society of America.vol. 69.pp. 293–297,1979.

[12] M.A.A.Nogol Sadat. "Seismotectonic map of Iran (scale 1:1 000 000)". Geological Survey of Iran, 1993.

[13] H.Ramazi. "Seismotectonic map of Iran", 1/2500000 scale, Engineering Geophysics (Ph.D. dissertation). Charles University, Prague, Czech Republic p. 100, 1995.

[14] Iranian National Committee on Large Dams (IRCOLD), "Seismotectonic province of Iran". Report No. 12, 50 pp, 1996.

[15] N.Mirzaei, M.Gao, Y.T.Chen. "Seismic source regionalization for seismic zoning of Iran: major seismotectonic provinces". J. Earthquake Prediction Res. Vol.7,PP. 465–495,1998.

[16] B.Tavakoli, M.Ghafory-Ashtiany. "Seismic hazard assessment of Iran". Annals of Geophysics. Vol.42,PP. 1013–1021, 1999.

[17] A.Zamani, M.Nedaei, R.Boostani. "Tectonic zoning of Iran based on selforganizing map". Journal of Applied Sciences. Vol.9,PP. 4099–4114, 2009

[18] M.Mojarab, H.Memarian, M.Zare, A.H.Morshedy, M.H.Pishahang "Modeling of the seismotectonic provinces of Iran using the self-organizing map algorithm" Computers & Geosciences, Vol. 67, PP. 150–162, 2013.

[19] G.Piatetski-Shaprio, W.J.Frawley. "Knowledge Discovery in Databases". AAAI Press/The MIT Press, Monlo Park, CA, pp. 525, 1991.

[20] L.Kaufman, P.J.Rousseeuw "Finding groups in data: an introduction to cluster analysis." JOHN WILEY & SONS, INC. Hoboken, New Jersey, 2005.

[21] H.Shah, J.Undercoffer, A.Joshi "Fuzzy Clustering for Intrusion Detection" Fuzzy Systems. FUZZ '03. The 12th IEEE

International Conference on , Vol. 2, PP. 1274-1278, 2003.

[22] J.Bezdek, L.Hall, L.Clarke. "Review of MR image segmentation techniques using pattern recognition", Med. Phys, Vol. 20, No. 4, PP.1033 -1048, 1993.

[23] J.Wang, X.Wu, C.Zhang. "Support vector machines based on K-means clustering for real-time business intelligence systems", Int. J. Bus. Intell. Data Mining, Vol. 1, No. 1, PP.54 -64, 2005.

[24] O.Zamir, O.Etzioni. "Grouper: a dynamic clustering interface to Web search results", Proceedings of the eighth international conference on World Wide Web, Toronto, Canada, 1999.

[25] N.P.Lin, C.Chang, H.Chueh, H.Chen, W.Hao, "A deflected Grid-based Algorithm for Clustering Analysis" W. Trans. On Comp. VOL. 7, Issue 4, PP.125-132, 2008.

[26] J.Han, M.Kamber, J.Pe. "Data Mining: Concepts and Techniques ".Morgan Kaufmann Publishers, USA. 2011.

[27] P.Berkhin. "Survey of clustering data mining techniques". Technical report, Accrue Software, San Jose, CA, pp.56, 2002.

[28] A.K.Jain, M.N.Murty, P.J.Flynn. "Data clustering: a review". ACM Computing Surveys ,vol.31,pp. 264–323, 1999.

[29] G.Gan, C. Ma, J. Wu. "Data clustering theory, algorithms, and applications", Society for Industrial and Applied Mathematics, Philadelphia, Pennsylvania, USA, 2007.

[30] C.A.Charu , K.R.Chandan. "Data Clustering: Algorithms and Applications " , Chapman and Hall/CRC, 6000 Broken Sound Parkway NW, Suite 300 Boca Raton, 2013.

[31] F.Hosseininezhad, A.Salajegheh "Study and Comparison of Partitioning Clustering Algorithms", Iranian Journal of Medical Informatics,Vol.2,Issue1, 2012.

[32] C.A.Charu. K.R.Chandan. "Data Clustering: Algorithms and Applications " , Chapman and Hall/CRC, 6000 Broken Sound Parkway NW, Suite 300 Boca Raton, 2013.

[33] M.Hung, D.Yang "An efficient fuzzy c-means clustering algorithm", Proc. IEEE Int. Conf. Data Mining, pp.225 -232, 2001.

[34] D.W.Kim, K.H.Lee, D.Lee. "On cluster validity index for estimation of the optimal number of fuzzy clusters". Pattern Recognit. Vol.37,pp. 2009–2025, 2004.

[35] A.Topchy, J.W.Punch. "Combining multiple weak clusterings". In: Proceedings of the Third IEEE International Conference on Data Mining (ICDM003), pp. 331–338, 2003.

[36] J.MacQueen. " Some methods for classification and analysis of multivariate observations " In Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Vol. 1, PP. 281 – 297, Berkeley, CA, USA, 1967.

[37] L.Kaufman, PJ.Rousseeuw ."Clustering Large Applications (Program CLARA). Finding groups in data" an introduction to cluster analysis,pp.126–163, 1990.

[38] M.Halkidi, Y.Batistakis, M.Vazirgiannis. "Clustering Validity Checking Methods: Part Ⅱ." ACM SIGMODRecord, Vol. 31 Issue 2, PP. 19-27, 2002.

[39] L.Davies, L.David, L.Bouldin, W.Donald. "A Cluster Separation Measure". IEEE Transactions on Pattern Analysis and Machine Intelligence. PP. 224–227, 1979.

[40] PJ.Rousseeuw . "Silhouettes: a Graphical Aid to the Interpretation and Validation of Cluster Analysis". Computational and Applied Mathematics,vol. 20,pp. 53–65, 1987.

[41] M.Halkidi, Y.Batistakis, M.Vazirgiannis. " Quality scheme assessment in the clustering process". In PKDD, London, UK,. Springer-Verlag.pp.s 265–276, 2000.

[42] M.Halkidi, Y.Batistakis, M.Vazirgiannis. " Clustering validity assessment: Finding the optimal

partitioning of a data set". In Proceedings of the 2001 IEEE International Conference on Data
Mining, pp. 187–194, 2001.

[43] R.Minho Kim, S.Ramakrishna. "New indices for cluster validity assessment". Pattern
Recognition Letters, vol.26,issue,15.pp.2353–2363, 2005.

[44] M.Shahvar, M.Zare. S.Castellaro " A unified seismic catalog for the Iranian plateau (1900–2011)" Geoscienceworld, Vol. 84 ,no. 2, pp. 233-249, 2013.

K. N. TOOSI UNIVERSITY OF TECHNOLOGY
FACULTY OF GEODESY AND GEOMATICS ENGINEERING

Journal of Geospatial Information Technology
Vol.3, No.1, spring 2015

Determination of the Number of Seismotectonic Provinces using Partitioning Clustering of Seismic Data

Mohammad Kazemi Beydokhti[1], Rahim Ali Abbaspour[*2]

1- MSc Student of GIS, School of Surveying and Geospatial Engineering, College of Engineering, University of Tehran.
2- Assistant Professor, School of Surveying and Geospatial Engineering, College of Engineering, University of Tehran.

Abstract

Seismotectonic provinces are defined as areas that have similar seismological characteristics, active tectonic fault system similar to each other. Identification of these areas is one of the most important subjects in seismic hazard analysis. Clustering methods are known as techniques for identifying information associated with the huge number of data. Since these methods are appropriate visual representations and provide reliable results, they can be an important tool for determining the seismotectonic provinces. In this study, three partitioning clustering methods were used including: K-means, CALARA, and Fuzzy C-means. For assessing the validity of these methods, five clustering validity indexes including: Davies-Bouldin, Silhouette, SD, S_Dbw, and Xie-Beni were used. Parameters including the earthquake epicenters, magnitude, gravity, magnetism, and fault density, as input clustering methods, were considered for modeling seismotectonic regions. After determining the optimal number of clusters, in order to evaluate the results of clustering model IRCOLD seismotectonic model was applied. The results showed that compared with other methods, CLARA have a greater compatibility with seismotectonic model in Iran. Finally, after comparing the number of optimized clustering methods with seismotectonic model it is concluded that the best number of seismotectonic provinces is 13.

Key words: Seismotectonic zones, clustering, validity indexes, seismic data

Correspondence : Address GIS Group, School of Surveying and Geospatial Engineering, College of Engineering, University of Tehran, Tehran, Iran.
Tel : 88008841
Email : abaspour@ut.ac.ir