

طبقه‌بندی غیرنظارتی داده‌های لرزه‌ای با استفاده از مدل‌های ترکیبی گوسی

امیر آریان نژاد^۱، محمدرداد*^۲ و سعید هادیلو^۳

۱- گروه نفت، دانشگاه آزاد اسلامی واحد الکترونیکی، تهران، ایران

۲- دانشکده مهندسی معدن، نفت و ژئوفیزیک، دانشگاه صنعتی شاهرود، ایران

۳- پژوهشکده علوم پایه کاربردی جهاد دانشگاهی، تهران، ایران

تاریخ دریافت: ۱۳۹۸/۰۶/۱۹ تاریخ پذیرش: ۱۳۹۹/۰۲/۱۶

چکیده

تحلیل رخساره‌های لرزه‌ای نقش بسزایی در مطالعات مخازن هیدروکربنی دارد. از آن‌جا که در ابتدای شروع عملیات اکتشافی مخازن هیدروکربنی تعداد چاه‌های موجود در منطقه کم است و یا وجود ندارد، می‌توان با استفاده از الگوریتم‌های شناسایی الگو و نشان‌گرهای مناسب لرزه‌ای به بررسی تغییرات جانبی و آنالیز رخساره‌های لرزه‌ای در یک افق پرداخت. روش‌های نظارتی و غیرنظارتی کمک شایانی به‌دقت، سرعت و کاهش هزینه دسته‌بندی داده‌ها و تحلیل بهتر رخساره‌های لرزه‌ای می‌کنند. اساس روش‌های غیرنظارتی، که موضوع این مقاله است، دسته‌بندی تمام داده‌ها در فضای نشان‌گری است و نتیجه به‌دست آمده به اطلاعات پیشین بستگی ندارد. در این روش، دسته‌بندی و تفسیر نتایج به‌دست آمده، بدون استفاده از داده‌های چاه، با آنالیز تطابق بین رخساره‌های لرزه‌ای به‌دست می‌آید. روش‌های گوناگونی برای خوشه‌بندی غیرنظارتی وجود دارد. در این مقاله روش مدل‌های ترکیبی گوسی (GMM) مورد استفاده قرار گرفته است که با استفاده از توزیع گوسی و اختصاص احتمال عضویت به نمونه‌های تحلیل، خوشه‌بندی آن‌ها را تعیین می‌کند. با استفاده از این روش تحلیل رخساره لرزه‌ای بر روی یک مجموعه داده سه‌بعدی مربوط به یکی از میدین هیدروکربنی جنوب ایران انجام می‌شود. تحلیل بر روی دو افق مختلف انجام شده است و نتایج نشان می‌دهد که روش GMM عملکرد قابل قبولی در خوشه‌بندی دارد و نتایج با تحلیل کیفیت مخزنی حاصل از رخساره‌های الکتریکی در برخی چاه‌ها هم‌خوانی دارد.

کلمات کلیدی: رخساره لرزه‌ای، خوشه‌بندی، طبقه‌بندی غیرنظارتی، مدل ترکیبی گوسی، کیفیت مخزنی

مقدمه

از چاه‌ها به‌دست می‌آید که شامل نمودار چاه‌ها، نمونه‌های مغزه و داده‌های تولید است. با این حال، نمودار چاه و داده مغزه، اندازه‌گیری‌های محلی هستند که ممکن است تمامی رفتار مخزن را منعکس نکنند. در مقابل داده‌های پراکنده چاه، داده لرزه‌ای سه‌بعدی منطقه بزرگی را پوشش می‌دهد.

انواع زیادی از اطلاعات در ساخت مدل مخزن استفاده می‌شوند. یکی از مهم‌ترین منابع اطلاعات،

*مسئول مکاتبات
آدرس الکترونیکی: mradad@shahroodut.ac.ir
شناسه دیجیتال: (DOI: 10.22078/pr.2020.3923.2786)

از روش k-mean برای خوشه‌بندی غیرنظارتی در یک مخزن کربناته در برزیل استفاده کردند.

روش دیگر در این زمینه ترسیم متقابل^۲ است. مفسران، مدت زیادی از بسته‌های نرم‌افزاری تجاری به‌منظور ترسیم متقابل نشان‌گرهای لرزه‌ای دوبعدی و سه‌بعدی در مخزن به‌منظور تعریف رخساره‌های سنگی استفاده می‌کردند. برای واسنجی چنین ترسیم‌هایی، مفسر، اطلاعات نمودار چاه را با استفاده از اشعه گاما و پاسخ مقاومت ویژه همراه با نگاره‌های موج P و موج S و چگالی، ترسیم متقابل می‌کند تا توابع چگالی احتمال را برای هر رخساره دلخواه بسازد و در نتیجه ترسیم‌های متقابل لرزه‌ای را واسنجی کند. با محصور کردن نقاط خوشه‌های ناهنجار در فضای ترسیم متقابل و تصویر کردن آن‌ها روی مقاطع افقی یا عمودی، رخساره‌های مورد نظر روی برش‌های افقی یا عمودی ترسیم می‌شوند. با این حال ترسیم متقابل بیش از سه نشان‌گر، نه فقط از منظر پیاده‌سازی، بلکه درک آن نیز دشوار می‌شود [۱۵]. تحلیل مؤلفه اصلی (PCA)^۳ تکنیک دیگری است که در خوشه‌بندی بدون نظارت به‌کار می‌رود و یک ابزار مفید کاهش بعد است. خیلی از نشان‌گرها از طریق زمین‌شناسی به هم جفت می‌شوند، به‌طوری‌که یک گسل می‌تواند باعث افزایش تغییرات جانبی در شکل موج، شیب، فرکانس و دامنه شود. خیلی از نشان‌گرها، مثل معیارهای گوناگون شباهت یا مجموعه‌ای از مؤلفه‌های طیفی نزدیک به هم، از طریق محاسبات ریاضی جفت می‌شوند [۱۶]. میزان فراوانی نشان‌گرها توسط ماتریس کواریانس اندازه‌گیری می‌شود. از نظر ریاضی، تعداد نشان‌گرهای مستقل خطی با مقادیر ویژه و بردارهای ویژه ماتریس کواریانس تعریف می‌شود.

تغییرات سنگ‌شناسی و سیالات، در تغییرات دامنه، شکل، همبستگی جانبی و نشان‌گرهای لرزه‌ای دیگر تأثیر می‌گذارند. نشان‌گرهای لرزه‌ای می‌توانند اطلاعات ساختار مدل مخزن را فراهم کنند. توصیف و تفسیر پارامترهای بازتابی لرزه‌ای استخراج شده که شامل هندسه، پیوستگی، دامنه، فرکانس و سرعت بازه‌ای می‌باشند، به‌عنوان تحلیل رخساره‌های لرزه‌ای شناخته می‌شوند [۱]. تحلیل رخساره‌های لرزه‌ای می‌تواند از طریق استفاده از تکنیک‌های شناسایی الگو انجام شود. با ترکیب مناسب نشان‌گرهای لرزه‌ای، می‌توان تغییرات جانبی مخزن را شناسایی و سپس با اطلاعات چاه واسنجی^۱ کرد [۲ و ۳]. وقتی که اطلاعات زمین‌شناسی، ناقص یا ناموجود باشد، تحلیل رخساره‌های لرزه‌ای را بدون نظارت می‌نامند و از طریق یادگیری غیرنظارتی یا الگوریتم‌های خوشه‌بندی اجرا می‌شود [۴]. استفاده از طبقه‌بندی رخساره لرزه‌ای با کمک کامپیوتر، بلافاصله بعد از توسعه نشان‌گرهای لرزه‌ای در دهه ۱۹۷۰ [۵ و ۶]، با تحقیقات سانلند [۷] و جاستایس و همکاران [۸] شروع شد. در عمل، یادگیری بدون نظارت از نشان‌گرها به‌عنوان داده یادگیری و داده تحت آنالیز استفاده می‌کند.

بارنز و لافلین [۹]، ژائو و همکاران [۱۰]، چوپرا و مارفورت [۱۱] و هادیلو و همکاران [۱۲] مقایسه‌ای بین روش‌های مختلف بدون نظارت برای طبقه‌بندی رخساره‌های لرزه‌ای انجام داده‌اند. ساده‌ترین الگوریتم خوشه‌بندی بدون نظارت، k-mean است [۱۳] که در آن مفسر تعداد رخساره‌هایی (خوشه‌هایی) را که باید پیدا شوند، تعریف می‌کند. سپس الگوریتم، میانگین‌ها و انحراف‌های استاندارد (به‌طور کلی ماتریس‌های کواریانس) را پیدا می‌کند تا مرکز و گسترش هر کدام از خوشه‌ها در فضای نشان‌گری چندبعدی تعیین شود. با استفاده از قوانین طبقه‌بندی بی‌زی، بردار نشان‌گر در یک نقطه داده شده، به خوشه‌ای که به آن نزدیک‌تر است اختصاص داده می‌شود. فریرا و همکاران [۱۴]

1. Calibration

2. Cross-plotting

3. Principal Component Analysis

یادگیری، تابع همسایگی و نرخ یادگیری اشاره کرد به طوری که این پارامترها وابسته به داده هستند [۲۰]. بیشاپ و همکاران [۲۰] یک روش جایگزین را برای نقشه‌های خودسازمانده توسعه دادند که بر این محدودیت‌ها فائق آیند. این روش، الگوریتم نقشه‌های توپوگرافی مولد (GTM)^۴ نامیده می‌شود و یک تکنیک کاهش ابعاد غیرخطی است که نمایش احتمالی بردارهای داده را در یک فضای مخفی فراهم می‌کند. روی [۲۱] و روی و همکاران [۲۲]، جزئیات این روش را تشریح کردند و کاربرد آن را برای به نقشه درآوردن رخساره‌های لزره‌ای حوزه وراکروز در مکزیک بیان کردند.

مدل‌های ترکیبی گوسی (GMM)^۵، مدل‌های پارامتری توزیع احتمال هستند که می‌توانند انعطاف و دقت بیشتری را در مدل‌سازی نسبت به الگوریتم‌های خوشه‌بندی سنتی فراهم کنند. لوبو و همکاران [۲۳] این تکنیک را به یک مجموعه از نمودارهای چاه که در منطقه Horseshoe Atoll در تگزاس غربی به دست آمده‌اند، به منظور تولید سنگ‌شناسی‌های مختلف اعمال کرده‌اند. این سنگ‌شناسی‌های GMM، سپس جهت واسنجی نتایج وارون‌سازی پیش از برانبارش لزره‌ای سه‌بعدی به منظور ساخت یک مدل خواص سنگ سه‌بعدی استفاده شده‌اند. همچنین در زمینه استفاده از روش GMM می‌توان به تحقیق هاردیستی و والت [۲۴] اشاره کرد که با استفاده از الگوریتم پیشینه‌سازی انتظار به بررسی و تحلیل رخساره‌های لزره‌ای به وسیله خوشه‌بندی داده‌های لزره‌ای پرداختند. در ایران نیز پژوهش‌های مختلفی در زمینه به‌کارگیری روش‌های تحلیل رخساره لزره‌ای انجام شده است. عدالت و سیاهکوهی [۲۵] از تحلیل رخساره لزره‌ای برای توصیف یکی از مخازن ایران به دو شیوه با نظارت و بدون نظارت بهره بردند.

اگر مقادیر ویژه از بزرگ‌ترین به کوچک‌ترین مرتب شوند اولین بردار ویژه متناظر با بیشترین مقدار ویژه، اولین مؤلفه اصلی خواهد بود که نشان‌دهنده بیشترین تغییرات در داده است. بردارهای ویژه متناظر با مقادیر ویژه بعدی مؤلفه‌های اصلی بعدی می‌باشند. مؤلفه‌های اصلی بر هم متعامد هستند. به دلیل این که نشان‌گرهای لزره‌ای از طریق زمین‌شناسی محدودیت باند موجک چشمه با هم همبستگی دارند، دو یا سه مؤلفه اصلی اول، تقریباً همیشه، بیان‌گر اکثریت تغییرپذیری داده می‌باشند. رودن و همکاران [۱۷] از ترکیب PCA و نقشه‌های خودسازمانده (SOM)^۱ برای بازساخت الگوی زمین‌شناسی از نشان‌گرهای لزره‌ای استفاده کردند. روش نقشه‌های خودسازمانده یک تکنیک دیگر غیرنظارتی برای تولید نقشه رخساره‌های لزره‌ای از طریق نشان‌گرهای لزره‌ای چندگانه است. در یک توزیع با N نشان‌گر در یک فضای داده N بعدی، سطحی که بهترین تطابق را با داده دارد توسط دو بردار ویژه اول ماتریس کواریانس تعیین می‌شود. سپس صفحه مورد نظر به صورت تکرار شونده به یک سطح دوبعدی تحت عنوان گوناگونی تغییر شکل داده می‌شود که تطابق بهتری را با داده دارد. بعد از هم‌گرایی، داده‌های N بعدی روی این سطح دوبعدی تصویر می‌شوند. گائو [۱۸] نقشه خودسازمانده یک بعدی را روی یک ماتریس هم‌رخداد سطح خاکستری (GLCM)^۲ باهدف به نقشه درآوردن رخساره‌های لزره‌ای منطقه دریایی آنگولا پیاده‌سازی کرد. کیان و همکاران [۱۹] برای خوشه‌بندی روش SOM را بر روی داده‌های پیش از برانبارش^۳ و با توسعه یک رهیافت داده محور برای استخراج ویژگی به نام ثبت خودکار هم‌آمیختگی عمیق به کار گرفتند. نقشه خودسازمانده کوهنن به عنوان پرکاربردترین روش خوشه‌بندی غیرنظارتی علی‌رغم سادگی پیاده‌سازی و نرخ نه چندان بالای محاسبات محدودیت‌هایی نیز دارد که می‌توان به عدم وجود مبنای تئوری برای انتخاب شعاع

1. Self-Organized Maps

2. Grey-Level Co-occurrence Matrix

3. Stacking

4. Generative Topographic Maps

5. Gaussian Mixture Models

شهرت خود را به دست آورد. می‌توان گفت هارتلی [۳۰]، اولین نفری بود که مفهوم EM را معرفی کرد. با این حال، دمپستر و همکاران [۳۱]، این روش را فرمول‌بندی و پیاده‌سازی کردند. در اصل، الگوریتم EM یک روش بهینه‌سازی تکرارشونده برای تخمین پارامتر است. دو مرحله در هر تکرار وجود دارد: (۱) با تخمین پارامترهای مدل ترکیبی گوسی، از تکرار قبلی، اطلاعات عضویت (یعنی احتمالی که هر داده از هر مدل گوسی نمونه‌برداری شده باشد) محاسبه می‌شود. این مرحله، به‌عنوان مرحله انتظار یا مرحله E شناخته می‌شود. (۲) با استفاده از اطلاعات عضویت تخمین زده شده و اندازه‌گیری‌ها، یک تخمین درست‌نمایی بیشینه از پارامترهای مدل ترکیبی گوسی در مرحله دوم به دست می‌آید. این مرحله به‌عنوان مرحله بیشینه‌سازی یا مرحله M شناخته می‌شود. این فرآیند دو مرحله‌ای در هر تکرار به هنگام می‌شود تا اینکه هم‌گرایی حاصل شود [۳۲].

مدل ترکیبی گوسی

در علم آمار، مدل ترکیبی گوسی، یک توزیع احتمال است که یک ترکیب محدب از چند توزیع احتمال گوسی است. فرض کنید مقدار تصادفی d نماینده یک نمونه از یک ترکیب از M مدل گوسی مشخص باشد، پس تابع چگالی احتمال (pdf) d با توجه به مدل ترکیبی گوسی به صورت زیر خواهد بود:

$$P(d|P, Q) = \sum_{j=1}^M p_j P(d|m_j, C_j) \quad (1)$$

که $\Pi = (\pi_1, \pi_2, \dots, \pi_M)$ سهم‌های ترکیب می‌باشند و ترکیب محدب نیازمند آن است که برای $j=1, 2, \dots, M$ ، $\pi_j \geq 0$ و $\sum_{j=1}^M \pi_j = 1$ باشد [۳۳]. $P(d|m_j, C_j)$ یک توزیع گوسی با میانگین μ_j و ماتریس کواریانس C_j است. در رابطه (۱)، $\Theta = (\theta_1, \theta_2, \dots, \theta_M)$ و $\theta_j = (\mu_j, C_j)$ است.

اگر $\Phi = (\Pi, \Theta)$ نشان‌دهنده پارامترهای مدل ترکیبی گوسی باشد، پس رابطه (۱) به صورت زیر خواهد بود:

جوکار و همکاران [۲۶] رخساره‌های لرزهای را از طریق آنالیز چند نشان‌گری و با روش‌های نظارتی و غیرنظارتی دسته‌بندی کردند. هادیلو و همکاران [۲۷] با استفاده از تحلیل زمان-بسامد و استخراج نشان‌گر خط بیشینه دامنه ضرایب تبدیل موجک، و استفاده از روش‌های SOM و k-mean در جهت بهبود تحلیل رخساره لرزهای پژوهشی انجام دادند. ثابتی و همکاران [۲۸] از روش خوشه‌بندی سلسله مراتبی برای تحلیل رخساره لرزهای، و از طریق آزمایش الگوریتم برروی داده دوبعدی و سه‌بعدی، استفاده کردند. باقری و همکاران [۲۹] تحلیل رخساره لرزهای را در فضای عدم تشابه و بدون استفاده از نشان‌گرهای لرزهای انجام دادند. در این مقاله از روش GMM برای تحلیل رخساره لرزهای از طریق خوشه‌بندی بدون نظارت در یک مجموعه داده لرزهای سه‌بعدی مربوط به یکی از میادین هیدروکربنی جنوب ایران استفاده شده است. در ادامه مقاله پس از ارائه تفصیلی از مبانی تئوری روش GMM، نتایج به دست آمده از تحلیل رخساره لرزهای ارائه شده است. در پایان، نتایج حاصل از این مقاله بیان شده است.

تئوری

از یک منظر روش‌های خوشه‌بندی به دو دسته خوشه‌بندی کلاسیک (سخت) و خوشه‌بندی فازی (نرم) دسته‌بندی می‌شوند. خوشه‌بندی کلاسیک بدین صورت است که هر نمونه داده فقط می‌تواند در یک خوشه قرار بگیرد یعنی احتمال قرار گرفتن در دو یا بیش از دو خوشه را ندارد. به عبارت دیگر، می‌توان گفت که احتمال قرار گرفتن یک داده در خوشه‌ای خاص، یا صفر است یا یک؛ ولی در خوشه‌بندی فازی، احتمال قرار گرفتن یک نمونه داده در بیش از یک خوشه وجود دارد. روش GMM یک تکنیک خوشه‌بندی نرم است. رهیافت این روش پیاده‌سازی الگوریتم بیشینه‌سازی انتظار (EM) برای مدل‌های ترکیبی گوسی است. الگوریتم EM، در روش‌های آماری تخمین پارامتر،

که z_i یک ویژگی گروه برای داده d_i است. اگر داده d_i به گروه z_i تعلق داشته باشد یعنی از مدل گوسی زام نمونه‌برداری شود، پس $z_i = j$.

به‌دست آوردن تخمین درست‌نمایی بیشینه پارامترهای مدل ترکیبی گوسی Φ با بیشینه‌سازی مستقیم تابع درست‌نمایی لگاریتمی داده - ناقص (رابطه (۳))، کار دشواری است. در عوض، الگوریتم EM، به‌صورت تکراری، تابع درست‌نمایی لگاریتمی داده - ناقص را با افزایش تابع Q (انتظار تابع درست‌نمایی لگاریتمی داده - کامل) در هر تکرار، بهبود می‌دهد. کانگ و همکاران [۳۳] نشان داده‌اند که اگر تابع Q در هر تکرار افزایش یابد، آن‌گاه تابع درست‌نمایی لگاریتمی داده ناقص $L(\Phi|D)$ هم افزایش خواهد یافت. می‌توان نشان داد که رابطه ۵ معادل است با [۳۳]:

$$Q(\Phi|\Phi_n) = \sum_{i=1}^n \sum_{j=1}^M h_i^j(\Phi_n) \log[\pi_j P(d_i|\delta_i^j = 1, \Phi)], \quad (6)$$

که اگر δ_i^j اگر $z_i = j$ برابر با یک و اگر $z_i \neq j$ برابر با صفر است. $h_i^j(\Phi_n)$ انتظار δ_i^j با داده و پارامترهای مدل ترکیبی گوسی (Φ_n) مشخص در تکرار n ام است، یعنی:

$$h_i^j(\Phi_n) = E_z\{\delta_i^j | D, \Phi_n\} = P(\delta_i^j = 1 | d_i, \Phi_n) = \frac{(\pi_j)_n P(d_i | (\mu_j)_n, (C_j)_n)}{\sum_{k=1}^M \{(\pi_k)_n P(d_i | (\mu_k)_n, (C_k)_n)\}} \quad (7)$$

که

$$P(d_i | (\mu_j)_n, (C_j)_n) = \frac{1}{(2\pi)^{\frac{N_d}{2}} |(C_j)_n|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(d_i - (\mu_j)_n)^T (C_j)_n^{-1} (d_i - (\mu_j)_n)\right] \quad (8)$$

در رابطه ۸، N_d بُعد بردار داده d_i و $|(C_j)_n|$ ، دترمینان ماتریس کواریانس $(C_j)_n$ هستند. در تکرار $(n+1)$ ام، مرحله E ماتریس عضویت را با استفاده از رابطه ۷ همراه با رابطه ۸ محاسبه می‌کند. از رابطه ۷، دریافت می‌شود که $h_i^j(\Phi_n)$ احتمال تعلق داده d_i به گروه z_i به گروه z_i تعیین می‌کند که اطلاعات گروه‌بندی داده می‌باشد (داده d_i از مدل گوسی زام نمونه‌برداری شده است). ماتریس عضویت $H = H(\Phi_n)$ در تکرار n ام، ماتریسی $N \times M$ می‌باشد که $h_i^j = H_{ij}(\Phi_n)$ به‌عنوان درایه سطر i ام و ستون j ام این ماتریس است. مرحله

$$P(d|\Phi) = \sum_{j=1}^M \pi_j P(d|\theta_j) \quad (2)$$

اگر فرض شود که داده‌های لرزه‌ای هر رخساره زمین‌شناسی از توزیع احتمال گوسی پیروی می‌کنند، بنابراین داده لرزه‌ای تمام مخزن که شامل ترکیب چند رخساره است، نمونه‌هایی از مدل ترکیبی گوسی می‌باشند. برای شناسایی رخساره‌های زمین‌شناسی از داده‌های لرزه‌ای نیاز است که مدل‌های گوسی و مدلی که از آن هر داده نمونه‌برداری شده باشد، پیدا شود [۳۲].

الگوریتم EM

اگر فرض شود که یک مجموعه از مقادیر اندازه‌گیری شده $D = (d_1, d_2, \dots, d_N)$ نمونه‌های یک مدل ترکیبی گوسی مشخص باشند رابطه ۲، تابع درست‌نمایی لگاریتمی پارامترهای مدل گوسی Φ با داده D به‌صورت زیر خواهد بود [۳۲]:

$$L(\Phi|D) = \log P(D|\Phi) \quad (3)$$

که \log بیان‌گر لگاریتم طبیعی است. اگر مشخص باشد که هر داده از کدام مدل گوسی نمونه‌برداری شده است، یعنی گروه‌بندی داده‌ها معلوم باشد، آن وقت می‌توان به‌راحتی تخمین درست‌نمایی بیشینه Φ را تعیین کرد. اما به‌طور کلی فقط داده وجود دارد و هیچ اطلاعاتی در مورد گروه‌بندی موجود نیست. در این مورد، تعیین پارامترهای مدل گوسی Φ یک مسئله داده - ناقص می‌شود. این مسئله داده - ناقص می‌تواند به‌صورت تکراری با استفاده از الگوریتم EM حل شود [۳۲]. تابع درست‌نمایی لگاریتمی به‌صورت زیر بیان می‌گردد [۳۱]:

$$L(\Phi|D) = Q(\Phi|\Phi_n) + R(\Phi|\Phi_n) \quad (4)$$

که Φ_n پارامترهای مدل گوسی از تکرار قبلی n ام می‌باشند و Q ، انتظار تابع درست‌نمایی داده - کامل است که به‌صورت زیر تعریف می‌شود:

$$Q(\Phi|\Phi_n) = E_z\{\log[P(D, Z|\Phi)] | D, \Phi_n\} \quad (5)$$

و R تفاضل بین درست‌نمایی لگاریتمی داده - ناقص (رابطه (۳)) و درست‌نمایی لگاریتمی داده - کامل (تابع Q) است. در رابطه ۵ $Z = (z_1, z_2, \dots, z_N)$ است

زیر تعریف می‌شود:

$$G(H) = \frac{1}{2} \sum_{k=1}^N \sum_{i=1}^N \sum_{j=1}^M h_i^j h_k^j v_{ik}, \quad (12)$$

که v_{ik} همسایگی را تعریف می‌کند، یعنی مقدار v_{ik} برای حالتی که d_i و d_k همسایه باشند و $i \neq k$ برابر با ۱ و در حالات دیگر برابر با صفر است [۳۲]. NEM در مقایسه با الگوریتم استاندارد EM در بخش درایه‌های ماتریس عضویت متفاوت است به طوری که درایه‌ها شامل اطلاعات فضایی نقطه داده جاری و همسایه‌های آن است. در این حالت خواهیم داشت:

$$(h_i^j)_n = \frac{(\pi_j)_n P(d_i | \mu_j, C_j)_n \exp(\beta \sum_{k=1}^N h_k^j v_{ik})}{\sum_{j=1}^M \{(\pi_j)_n P(d_i | \mu_j, C_j)_n \exp(\beta \sum_{k=1}^N h_k^j v_{ik})\}}, \quad (13)$$

که عبارت فضایی $\exp(\beta \sum_{k=1}^N h_k^j v_{ik})$ به‌عنوان یک عبارت نرم‌کننده فضایی روی ماتریس عضویت عمل می‌کند. برای محاسبه $G(H)$ ، از h_k^j محاسبه شده در تکرار قبلی سمت راست رابطه ۱۳ استفاده می‌شود [۳۲]. به‌دلیل این که عبارت فضایی در تابع راست‌نمایی لگاریتمی افزوده در رابطه ۱۱ شامل پارامترهای مدل ترکیبی Φ نیست، بیشینه‌سازی نسبت به Φ ، همانند الگوریتم EM استاندارد است. به‌هنگام‌سازی Φ هم همانند رابطه ۱۰ می‌باشد.

پیاده‌سازی روش GMM بر روی داده‌ها

در این بخش از مقاله در ابتدا برای بررسی عملکرد روش خوشه‌بندی GMM یک مجموعه داده مصنوعی مورد ارزیابی قرار می‌گیرد و پس از بررسی عملکرد روش در خوشه‌بندی این مجموعه داده، داده‌های لرزهای مورد تحلیل قرار می‌گیرند.

اعمال روش GMM روی داده‌های مصنوعی

برای ارزیابی کارایی روش GMM از داده‌های نمونه گل زنبق که حاصل اندازه‌گیری ویژگی‌های گلبرگ‌های ۱۲۰ نمونه گل از سه گروه مختلف (virginica و setosa، versicolor) است استفاده می‌شود.

اول در الگوریتم EM، محاسبه درایه‌های ماتریس عضویت h_i^j برای Φ_n مشخص از رابطه ۷ می‌باشد. همان‌طور که h_i^j ، انتظار δ_i^j می‌باشد، این مرحله، مرحله انتظار نامیده می‌شود. در مرحله بیشینه‌سازی (مرحله M) تکرار $(n+1)$ ام، باید پارامترهای مدل ترکیبی گوسی Φ_{n+1} از طریق بیشینه‌سازی تابع Q، که در رابطه ۶ تعریف شده، با ماتریس عضویت محاسبه شده از رابطه ۷ و در مرحله انتظار (مرحله E) تخمین زده شود؛ یعنی:

$$\Phi_{n+1} = \arg \max_{\Phi} Q(\Phi | \Phi_n). \quad (9)$$

حل‌های مرحله بیشینه‌سازی به‌صورت زیر خواهد بود:

$$\begin{aligned} (\mu_j)_{n+1} &= \frac{\sum_{i=1}^N (h_i^j)_n d_i}{\sum_{i=1}^N (h_i^j)_n}, \\ (C_j)_{n+1} &= \frac{\sum_{i=1}^N (h_i^j)_n (d_i - (\mu_j)_{n+1})(d_i - (\mu_j)_{n+1})^T}{\sum_{i=1}^N (h_i^j)_n}, \end{aligned} \quad (10)$$

$$(\pi_j)_{n+1} = \frac{1}{N} \sum_{i=1}^N (h_i^j)_n,$$

که $(h_i^j)_n = h_i^j(\Phi_n)$. اگر هم‌گرایی صورت نگرفته باشد، این حل‌ها برای شروع تکرار بعدی استفاده می‌شوند.

الگوریتم EM همسایگی (NEM) ۱

الگوریتم NEM، پیوستگی فضایی داده‌ها را در هر گروه با اضافه کردن یک عبارت جریمه فضایی به تابع درست‌نمایی لگاریتمی داده- ناقص اصلی، بهبود می‌دهد [۳۴] تا U این‌گونه به‌دست آید:

$$U(\Phi | D) = L(\Phi | D) + \beta G(H), \quad (11)$$

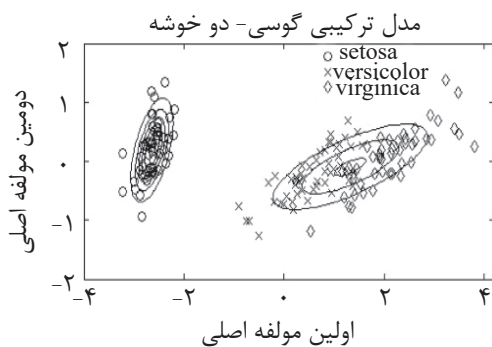
که β در عبارت جریمه، یک پارامتر قابل تنظیم است و با آزمایش تعیین می‌شود. وقتی β خیلی کوچک باشد، نتایج الگوریتم NEM مشابه نتایج الگوریتم EM استاندارد می‌شوند. وقتی β خیلی بزرگ باشد، NEM تأکید بیش از حد بر پیوستگی فضایی یک رخساره دارد و منجر به تعداد کمی رخساره می‌شود. مطالعه حساسیت توسط همان [۳۵] نشان داد که مقدار ۰/۰۱ تخمین‌های منطقی از نقشه‌های رخساره برای همه نمونه‌های مورد نظر ارائه می‌دهد. تابع قید فضایی $H(G)$ به‌صورت

خوشه‌بندی می‌شود و کارآیی این روش برای تحلیل رخساره‌های لرزه‌ای بررسی می‌گردد. تعیین محدوده‌های مستعد دارای کیفیت مخزنی مطلوب در اکثر مطالعات اکتشافی از اهمیت بالایی برخوردار است و به‌طور کل تشخیص بخش‌های مخزنی که وابسته به خواص اصلی سنگ‌ها است، نقش مهمی را در مطالعات نفتی ایفا می‌کند. برای دستیابی به این دانش اولین اطلاعات در اختیار، مغزه‌ها (که عموماً تنها در برخی از چاه‌های منطقه تولید می‌شوند) و نگارهای چاه می‌باشند، که اطلاعات دقیقی از ویژگی‌های سنگ‌شناسی و فیزیکی از محل چاه‌ها در اختیار قرار می‌دهند؛ ولی این اطلاعات با توجه به ماهیت چاه تنها مربوط به محل چاه است و فضای بین چاه‌های حفاری شده را شامل نمی‌شود. برای بررسی ویژگی‌های فیزیکی سنگ‌ها در فضای بین چاه‌ها اطلاعات لرزه‌ای در اختیار است تا با استفاده از ابزار ریاضی ویژگی‌های فیزیکی لایه‌ها و واحدهای زمین‌شناسی بررسی شود.

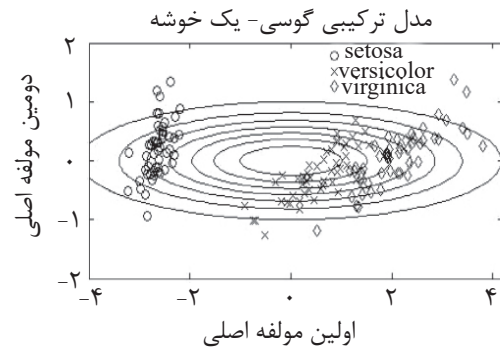
در این ارزیابی همچنین از روش PCA استفاده می‌شود تا با تولید دو ویژگی از چهار ویژگی اولیه امکان نمایش بهتر تغییرات در فضای دوبعدی را داشته باشیم. برای بررسی کارآیی روش GMM در این مثال از تعداد مختلف خوشه استفاده می‌شود تا نشان داده شود که در روش GMM به تعداد خوشه انتخابی، تابع توزیع گوسی اولیه تولید می‌گردد و سپس مراحل مختلف با بیشینه‌سازی تابع Q بهترین تابع و یا توابع گوسی را روی اطلاعات به‌دست می‌آورد. برای داده‌های نمونه با سه تعداد خوشه مختلف (تعداد خوشه یک تا سه)، خوشه‌بندی با استفاده از روش GMM انجام شده است که نتایج آن در شکل‌های ۱ تا ۳ نمایش داده شده است. همان‌گونه که مشاهده می‌شود روش GMM با استفاده از سه خوشه توانسته است سه نوع مختلف گل را تفکیک کند.

پیاپی‌سازی روش بر روی داده‌های لرزه‌ای

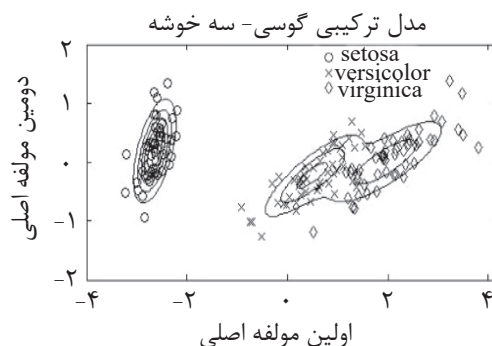
در این بخش اطلاعات لرزه‌ای سه‌بعدی در یک میدان نفتی در جنوب ایران توسط روش GMM



شکل ۲ خوشه‌بندی دوبعدی بر مبنای تعداد خوشه (۲ خوشه)



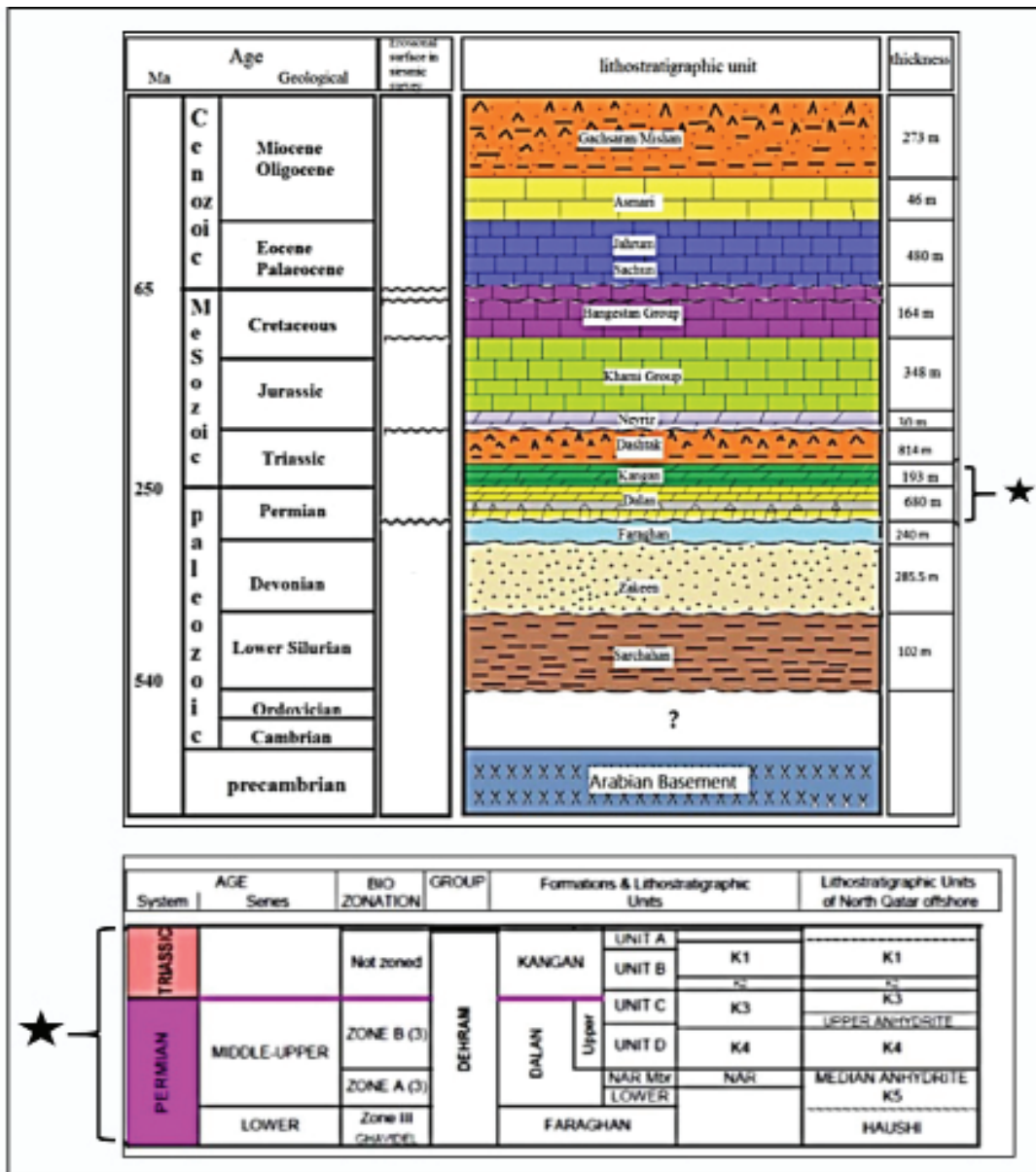
شکل ۱ خوشه‌بندی دوبعدی بر مبنای تعداد خوشه (۱ خوشه)



شکل ۳ خوشه‌بندی دوبعدی بر مبنای تعداد خوشه (۳ خوشه)

دالان بالایی شامل مجموعه‌ای از ریف‌های کربناته-تبخیری و کنگان با لیتولوژی غالب آهک و دولومیت با لایه‌هایی از انیدریت، توالی پرمو-تریاس را تشکیل می‌دهند. توالی دالان بالایی-کنگان در این میدان، شامل چهار واحد مخزنی است که از پایین به بالا شامل واحدهای K4، K3، (دالان بالایی)، K2 و K1 (کنگان) است. برای بررسی کارآیی روش GMM از داده‌های لرزه‌ای سه‌بعدی دو افق متوالی K1 و K2 انتخاب شده و بر روی هر دو افق تحلیل رخساره لرزه‌ای انجام می‌گردد.

برای انجام این کار ابتدا با استفاده از اطلاعات چک شات، VSP و یا مدل سرعت در منطقه داده‌های چاه و اطلاعات لرزه‌ای از نظر تبدیل مقیاس زمانی به عمقی با یکدیگر همگام‌سازی می‌شود (در اینجا از اطلاعات چک‌شات استفاده شده است)، و در مرحله بعد ارتباط بین اطلاعات چاه و داده‌های لرزه‌ای در محل چاه بررسی شده و این ارتباط به دیگر مناطق که اطلاعات چاه را ندارد بسط داده می‌شود. در شکل ۴ توالی رسوبی و چینه‌ای محدوده مطالعاتی در میدان مورد نظر آورده شده است.



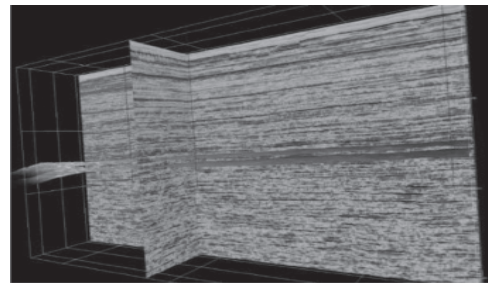
شکل ۴ نمایش چینه‌نگاری کلی به همراه واحدهای مخزنی [۳۶]

که دو افق کنگان و K4 بر روی آن مشخص هستند. داده‌های لرزه‌ای افق‌های K1 و K2 که مورد نظر این پژوهش هستند در شکل ۶ نمایش داده شده است. برای استخراج اطلاعات لرزه‌ای در روی افق‌های مورد نظر، از نتیجه دستچین^۱ هر افق استفاده می‌شود تا بر روی آن افق اطلاعات لرزه‌ای و نشان‌گرهای لرزه‌ای استخراج شود.

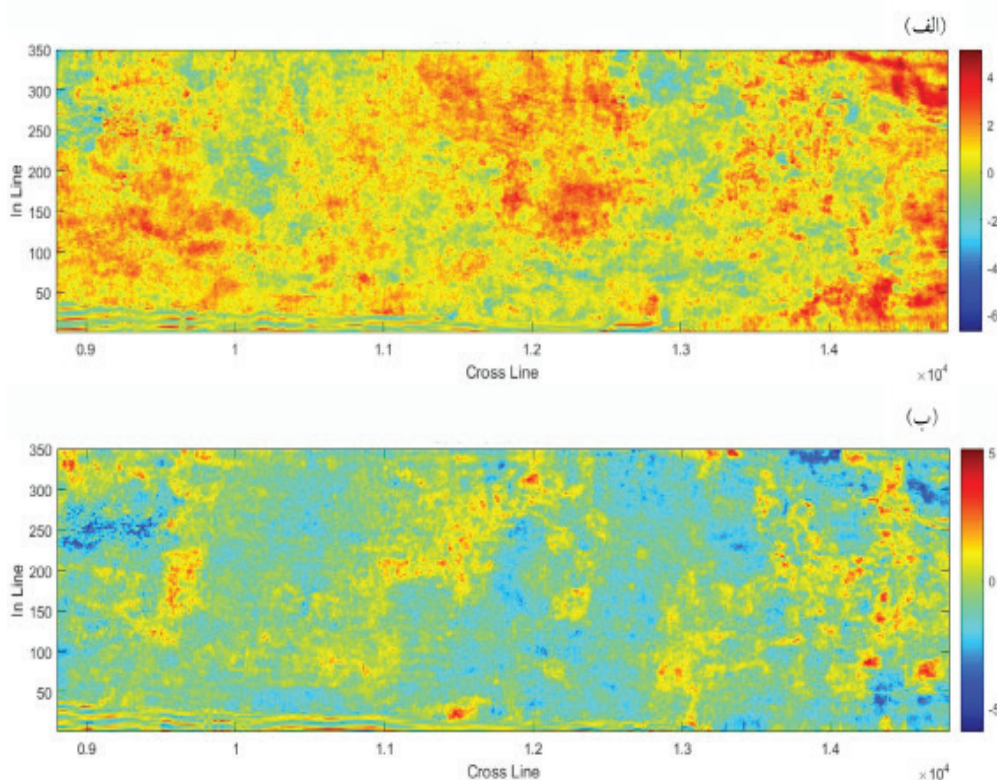
انتخاب نشان‌گرها

به منظور انجام یک خوشه‌بندی مناسب، باید ترکیب بهینه‌ای از نشان‌گرهای لرزه‌ای انتخاب و استفاده شود. در انتخاب نشان‌گرهای بهینه باید به این نکته توجه شود که هرکدام از آن‌ها بخشی از اطلاعات مفید لرزه‌ای را در اختیار داشته باشند و نمایش دهند. در این مسیر باید تلاش شود که تا حد ممکن از انتخاب نشان‌گرهایی که شباهت بالایی به یکدیگر دارند پرهیز شود [۱۶].

در تحلیل رخساره با استفاده از روش GMM توزیع رخساره‌های لرزه‌ای به دست می‌آید و همچنین میزان احتمال تعلق هر عضو به رخساره خود با عددی بین صفر تا یک تعیین می‌گردد که با استفاده از این توزیع احتمال کارآیی و صحت خوشه‌بندی قابل بررسی می‌شود. داده‌های لرزه‌ای در محدوده مطالعاتی، داده‌های سه‌بعدی بعد از برابارش مهاجرت داده شده هستند و در شکل ۵ در محیط نرم‌افزار OpendTect نمایش داده شده است.



شکل ۵ نمایشی از اطلاعات سه‌بعدی داده‌های لرزه‌ای در بین دو افق کنگان بالایی و K4 مورد استفاده در مطالعه



شکل ۶ نمایش اطلاعات دامنه لرزه‌ای نمونه اول زمانی افق (الف) K1 و (ب) K2

انجام تحلیل رخساره با استفاده از روش GMM برای خوشه‌بندی اطلاعات لرزه‌ای با استفاده از روش GMM باید تعداد خوشه و یا تعداد رخساره‌های لرزه‌ای برای انجام تحلیل رخساره لرزه‌ای مشخص باشد که برای این کار می‌توان از اطلاعات قبلی موجود در منطقه استفاده کرد و یا با استفاده از روش‌های ریاضی تعیین تعداد خوشه بهینه، تعداد مناسب خوشه را برای اطلاعات مورد نظر به دست آورد [۳۷ و ۳۸]. در این مطالعه براساس نتایج حاصل از مطالعه هادپلو و همکاران [۳۸] که در آن با استفاده از اطلاعات نگار سرعت امواج صوتی (DT)^۱، نگار تخلخل نوترونی (NPHI)^۲ و نگار چگالی (RHOB)^۳ و الکتروفاسیس‌های استخراج شده، کیفیت مخزنی در چاه‌های اکتشافی دسته‌بندی شد که اطلاعات آن در جدول ۱ آمده است. در همان مطالعه همچنین با استفاده از ارزیابی نتایج خوشه‌بندی بدون نظارت تعداد پنج خوشه به‌عنوان تعداد بهینه خوشه برای افق‌های مورد مطالعه شناسایی شد و بر همین اساس در این مطالعه از تعداد پنج خوشه برای تعیین گستره تغییرات خوشه‌ها در مطالعه بدون نظارت استفاده می‌شود [۳۹].

در این مطالعه از نشان‌گرهایی مانند دامنه لحظه‌ای، فرکانس لحظه‌ای، فاز لحظه‌ای و فرکانس غالب که اطلاعات پایه‌ای از داده‌های لرزه‌ای را دارند، و نشان‌گر بافتی میانگین GLCM که تغییرات احتمالی بافتی در افق‌ها را نمایندگی می‌کند، و همچنین نشان‌گرهای تک فرکانس ۲۰، ۳۰ و ۴۰ Hz حاصل از تجزیه طیفی که اطلاعات تغییرات فرکانس را در اختیار دارند و همچنین نشان‌گرهای انرژی و نشان‌گر تباين صوتی نسبی^۱ که اطلاعات مرتبط با امپدانس صوتی را دارد استفاده می‌شود. برای تحلیل رخساره لرزه‌ای در ابتدا هر کدام از نشان‌گرهای ذکر شده تولید می‌گردد. سپس با استفاده از ترسیم متقابل نشان‌گرهایی که در بیان تغییرات در داده شباهت عملکرد زیادی به هم دارند، برای اجتناب از افزودگی، مشخص و نشان‌گرهای اضافی حذف می‌شوند. سپس با استفاده از روش تحلیل مؤلفه اصلی از نشان‌گرهای باقی‌مانده و به‌منظور کاهش بُعد، متغیرهای (نشان‌گرهای) جدید تولید می‌شوند. بعد از استخراج اطلاعات نشان‌گرها روی افق مورد نظر، روش GMM برای خوشه‌بندی و تولید نقشه رخساره لرزه‌ای روی آن اجرا می‌شود.

جدول ۱ آنالیز نگارهای DT، NPHI و RHOB چاه‌های اکتشافی در منطقه و تفکیک خوشه‌ها براساس کیفیت مخزنی استنباط شده از نگارهای چاه [۳۹]

کد رخساره	DT	NPHI	RHOB	کیفیت مخزنی کدهای رخساره
۱	۴۸/۱	۰/۰۴	۲/۸۵	کیفیت غیر مخزنی
۲	۴۸/۳	۰/۰۵۱	۲/۸۸	کیفیت غیر مخزنی
۳	۵۰/۱	۰/۱۲۱	۲/۷۰۱	کیفیت مخزنی نه چندان بد
۴	۵۷	۰/۰۶۹	۲/۴	اساساً کلاس ۳ و تا حدودی کلاس ۲ با کیفیت مخزنی خوب
۵	۷۰	۰/۱۱۲	۱/۸	کلاس ۱ و ۲ با بهترین کیفیت مخزنی
۶	۵۶/۳	۰/۱۵۳۲	۲/۶	توزیع وسیع با کیفیت مخزنی متوسط تا خوب

1. Relative Acoustic Impedance

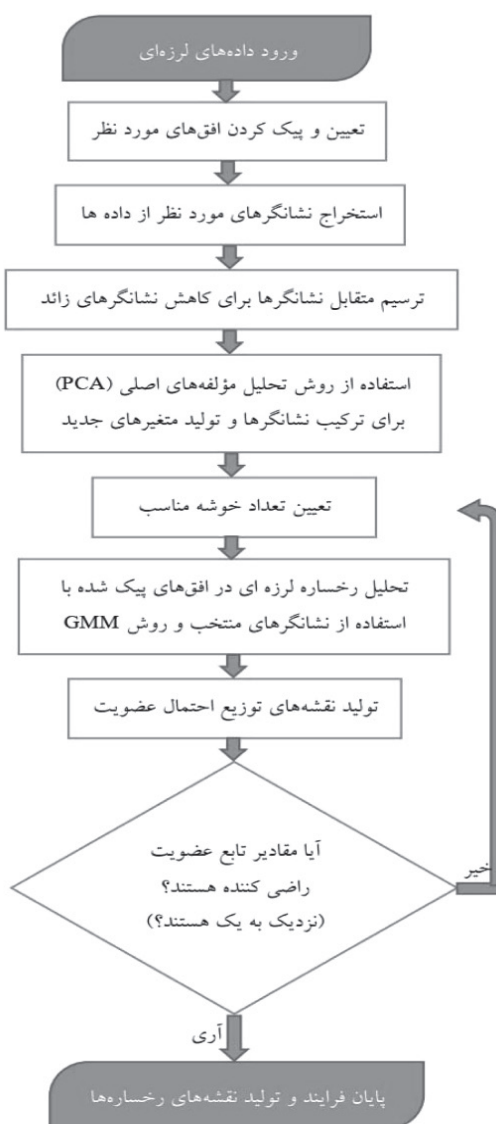
2. Acoustic Log

3. Neutron Porosity Log

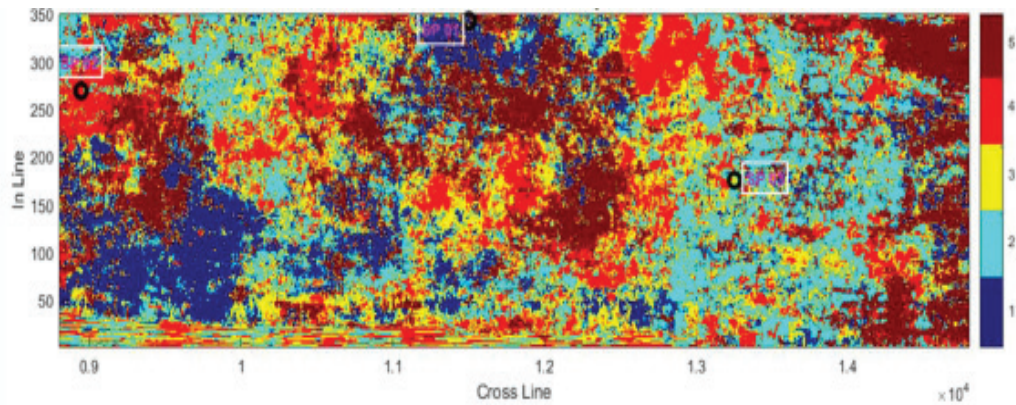
4. Density Log

می‌دهد. از بیان روابط در فلوجارت اجتناب شده است که پیش‌تر در بخش تئوری مقاله به تفصیل تشریح شده‌اند. نقشه‌های تحلیل رخساره حاصل از خوشه‌بندی اطلاعات لرزه‌ای در دو افق مورد نظر با استفاده از روش GMM و با در نظر گرفتن پنج خوشه تهیه شده است. شکل ۸ نتیجه تحلیل برروی افق K1 را نشان می‌دهد که برای این نقشه رخساره لرزه‌ای، نقشه توزیع احتمال عضویت نیز در شکل ۹ ارائه شده است. نقشه‌های تحلیل رخساره و نقشه توزیع احتمال عضویت متناظر برای افق K2 در شکل‌های ۱۰ و ۱۱ مشاهده می‌شود.

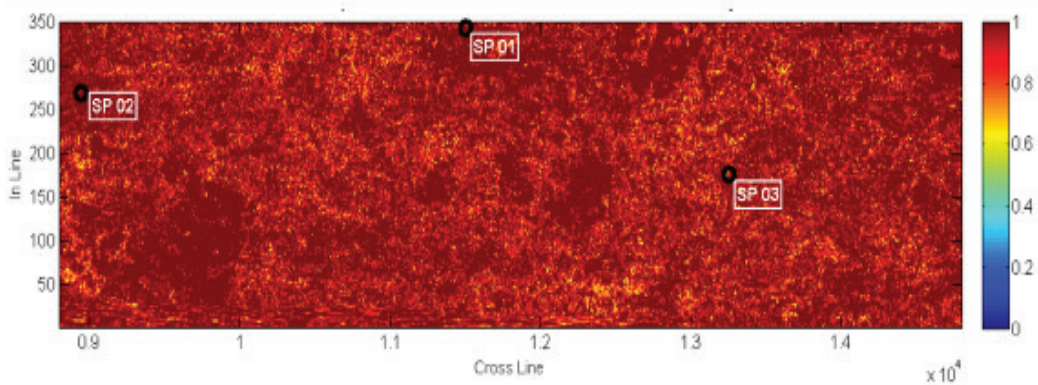
با توجه به اینکه روش GMM توزیع احتمال عضویت در خوشه را محاسبه می‌کند می‌توان از این مقادیر برای بررسی کلی بهینه بودن خوشه‌بندی استفاده کرد؛ به طوری که اگر برای بیشتر اعضا میزان عضویت عدد بالا و نزدیک به یک را نشان دهد می‌تواند گواه آن باشد که این تعداد خوشه توانسته داده‌ها را با دقت مناسبی تفکیک کند و اگر این چنین نبود می‌توان با تعیین تعداد خوشه جدید فرآیند را تکرار کرد. در این پژوهش تعداد پنج خوشه مناسب تشخیص داده شد. شکل ۷ فرآیند انجام تحلیل رخساره لرزه‌ای با روش GMM را نشان



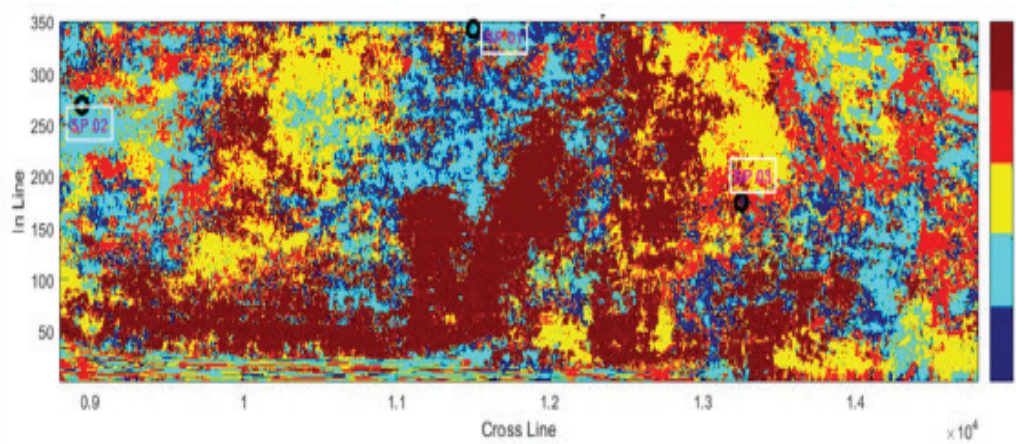
شکل ۷ فلوجارت فرآیند تحلیل رخساره لرزه‌ای با روش GMM



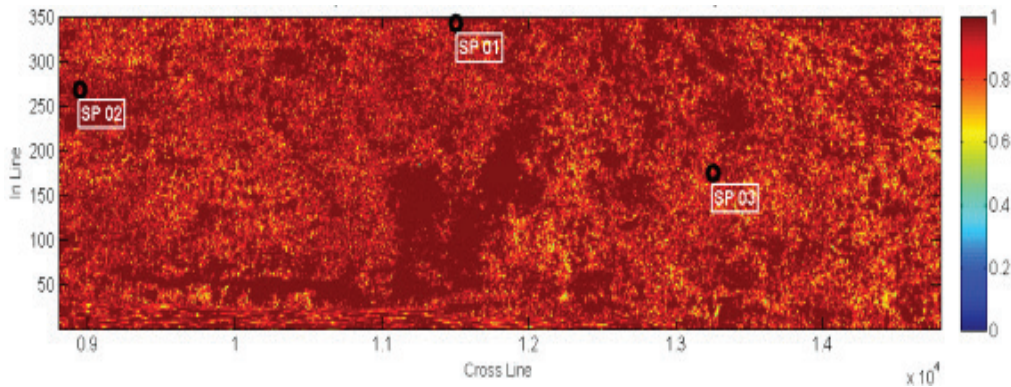
شکل ۸ نمایش اطلاعات لرزه‌ای پس از خوشه‌بندی با روش GMM (۵ خوشه) در روی افق K1. موقعیت سه چاه SP01، SP02 و SP03 با دایره‌های سیاه رنگ در شکل مشخص شده‌اند.



شکل ۹ نمایش توزیع احتمال عضویت در خوشه مربوطه در افق K1 که با روش GMM (۵ خوشه) خوشه‌بندی شده است.



شکل ۱۰ نمایش اطلاعات لرزه‌ای پس از خوشه‌بندی با روش GMM (۵ خوشه) در روی افق K2. موقعیت سه چاه SP01، SP02 و SP03 با دایره‌های سیاه رنگ در شکل مشخص شده‌اند.



شکل ۱۱ نمایش توزیع احتمال عضویت در خوشه مربوطه در افق K2 که با روش GMM (۵ خوشه) خوشه‌بندی شده است

چگالی در نگارهای چاه دارد (جدول ۱). با در اختیار بودن این ارتباط، اطلاعات نگارهای چاه توسط ابزار خوشه‌بندی بدون نظارت دسته‌بندی شد تا تفکیک رخساره‌ها برحسب نگارهای چاه به‌دست آید و با در نظر گرفتن خصوصیات مغزه‌های موجود کیفیت مخزنی دسته‌بندی نگارهای چاه تعیین شد (جدول ۱). با استناد به دسته‌بندی نگارهای چاه و خصوصیات مرتبط با آنها روش خوشه‌بندی بدون نظارت و با نظارت امکان پیش‌بینی رخساره‌ها را در کل منطقه‌ای که اطلاعات لرزه‌ای در آن برداشت شده ایجاد می‌کند. در این مطالعه روش بدون نظارت مدل‌های ترکیبی گوسی استفاده شده است تا با استناد به محل چاه بتوان ویژگی‌های دسته‌بندی‌های حاصل از اطلاعات لرزه‌ای را برآورد و بررسی کرد. برای ارزیابی نتایج خوشه‌بندی حاصل از روش GMM از رخساره کیفیت مخزنی تولید شده در مطالعه [۳۹] استفاده شد که اطلاعات آن برای دو افق K1 و K2 در جدول ۲ آورده شده است. طبق این اطلاعات چاه شماره دو و سه به ترتیب در خوشه پنج و شش رخساره کیفیت مخزنی قرار گرفته‌اند و بیان‌گر آن هستند که خوشه قرمز و آبی فیروزه‌ای در شکل ۸ کیفیت مخزنی بالاتری دارند ولی چاه شماره یک که در خوشه آبی پررنگ قرار گرفته با دارا بودن رخساره کیفیت مخزنی ۳، کیفیت مخزنی کمتری دارد.

در نقشه‌های توزیع احتمال عضویت هرچه عدد احتمال حضور به یک نزدیک‌تر باشد نشان از آن دارد که آن داده با احتمال بیشتر عضو خوشه مورد نظر است و آن داده در نزدیک مرکز توزیع گوسی قرار گرفته است. این نقشه به‌نحو مناسبی می‌تواند برای تعیین صحت و توانایی روش خوشه‌بندی استفاده شود. درحالی‌که همه اعضا احتمال حضور بالا و نزدیک به یک داشته باشند (رنگ قرمز تیره) می‌توان عنوان کرد که نشان‌گرها و تعداد خوشه‌ها به‌طور مناسب انتخاب شده‌اند. برای هر دو افق به‌صورت کیفی دیده می‌شود که توزیع احتمال عضویت اعدادی نزدیک به یک را به‌خود گرفته‌اند و این می‌تواند گواهی باشد از آن که خوشه‌بندی GMM توانسته با پنج خوشه دسته‌بندی مناسبی از اطلاعات لرزه‌ای را به‌دست دهد. در منطقه مورد مطالعه سه چاه وجود دارد که دو چاه (SP01 و SP02) اطلاعات مغزه دارند و دیگری تنها اطلاعات نگار در اختیار دارد. در مطالعه هادیلو و همکاران [۳۹] از این اطلاعات استفاده شد تا ارتباط بین مغزه‌ها و نگارهای چاه شناسایی و دسته‌بندی رخساره‌های مخزنی برحسب آنها انجام گردد. آنها ارتباط بین ویژگی‌های مغزه‌ها و نگارهای چاه را بررسی کردند و به‌طور مثال مشاهده کردند که مغزه‌های با درصد بالاتری از ماسه‌سنگ و یا دولومیت و درصد پایینی از انیدریت، ارتباط مستقیمی با بالا بودن میزان تخلخل و ارتباط معکوس با

رخساره مناسبی از اطلاعات به دست دهد. با در اختیار داشتن اطلاعات چاه و تطابق دادن اطلاعات چاه با رخساره‌های تولید شده می‌توان خصوصیات کامل رخساره‌ها را شناسایی کرد. الگوریتمی که در این مطالعه استفاده شد، تحلیل غیرنظارتی رخساره لرزه‌ای می‌باشد و این مطالعه بیشتر برای مراحل ابتدایی اکتشاف مورد استفاده قرار می‌گیرد و می‌تواند دید مناسب از گسترش رخساره‌های لرزه‌ای ارائه کند.

نتیجه‌گیری

در این مطالعه کاربرد روش GMM برای تحلیل رخساره غیر نظارتی لرزه‌ای استفاده شد. روش خوشه‌بندی GMM قادر است بدون استفاده از اطلاعات چاه نظم درون اطلاعات و توزیع رخساره لرزه‌ای را شناسایی کند. در این روش توابع گوسی با استفاده از بیشینه کردن تابع Q بهینه‌ترین حالت نمایندگی داده‌ها را پیدا کرده و داده‌ها را با توزیع احتمال مشخص در خوشه‌های مختلف قرار می‌دهد. روش GMM برای داده‌های گل که در آن با استفاده از روش PCA دو نشان‌گر جدید تولید شده بود توانست به خوبی سه خوشه موجود در اطلاعات را استخراج کند. در روش GMM از توابع گوسی به تعداد خوشه‌های تعیین شده برای تقریب و پوشش اطلاعات استفاده می‌شود. پس با توجه به اینکه داده‌های لرزه‌ای از توزیع گوسی پیروی می‌کنند انتظار حصول نتیجه از این روش برای خوشه‌بندی داده‌های لرزه‌ای منطقی می‌نماید. روش GMM برای اطلاعات لرزه‌ای سه‌بعدی در دو افق مختلف توزیع رخساره‌های لرزه‌ای را با همبستگی مناسبی بین افق‌ها تولید کرد و همچنین نشان داده شد که این روش با در اختیار قرار دادن نقشه توزیع احتمال عضویت ابزاری را در اختیار مفسر قرار می‌دهد تا بتواند در مورد کیفیت خوشه‌بندی انجام شده به صورت کیفی اظهار نظر کند. نتیجه خوشه‌بندی با تحلیل کیفیت مخزنی حاصل از رخساره‌های الکتریکی در چاه‌ها هم‌خوانی داشت.

جدول ۲ کیفیت مخزنی در محل چاه‌های موجود در منطقه و در محل پیک افق‌های K1 و K2 [۳۹]

چاه \ افق	SP01	SP02	SP03
K1	۳	۵	۶
K2	۳	۳	۴

برای افق K2 چاه شماره سه رخساره کیفیت مخزنی ۴ را دارد ولی دو چاه دیگر رخساره کیفیت مخزنی سه را دارا هستند و این نتیجه در خوشه‌بندی بدون نظارت نیز تکرار شده است و دو چاه یک و دو که رخساره کیفیت مخزنی یکسان داشتند در یک خوشه با رنگ آبی فیروزه‌ای قرار گرفته‌اند و چاه شماره ۳ با رخساره کیفیت مخزنی متفاوت در خوشه متفاوت قرار گرفته است (شکل ۱۰). دقت شود که خوشه‌های منفک شده در آنالیز رخساره (شکل‌های ۸ و ۱۰) که به تعداد ۵ عدد هستند، متفاوت از کد رخساره زمین‌شناسی (اعداد ۱ تا ۶ در جدول ۱) هستند. به این معنی که اگر به‌عنوان مثال ناحیه‌ای از شکل‌های ۸ یا ۱۰، با توجه به رنگ، مثلاً در خوشه ۵ قرار گرفت، لزوماً به این معنی نیست که آن ناحیه دارای کیفیت مخزنی متناظر با کد رخساره ۵ از جدول ۱ است. در واقع در این آنالیز می‌توان تشخیص داد که ناحیه‌ای از نقشه رخساره لرزه‌ای یک افق (شکل‌های ۸ و ۱۰) که دربردارنده یکی از چاه‌های SP01، SP02 و SP03 هستند، اگر در خوشه‌ای با رنگ مختص به خود قرار گرفت، بر مبنای اطلاعات چاه موجود در جدول ۲ مشاهده شود که این ناحیه متناظر با چه کد رخساره‌ای است؛ سپس به سراغ جدول ۱ رفته و بررسی شود که این کد رخساره متناظر با چه کیفیت مخزنی می‌باشد. در نتیجه می‌توان استنباط کرد در هر جای دیگر نقشه رخساره لرزه‌ای افق مورد نظر (شکل‌های ۸ و ۱۰) که رنگ مشابه وجود داشته باشد، معادل با کیفیت مخزنی مشابه است.

در این مطالعه تمرکز بر روی کارایی روش GMM بود تا نشان داده شود که این روش قادر است تحلیل

مراجع

- [1]. Mitchum RM (1988) Glossary of terms used in seismic stratigraphy, AAPG Memoir 26, 205–212: 1977.
- [2]. Dumay J, Fournier F (2005) Multivariate statistical analyses applied to seismic facies recognition, *Geophysics*, 60, 1437-1450.
- [3]. Saggaf MM, Toksöz MN, Marhoon MI (2003) Seismic facies classification and identification by competitive neural networks, *Geophysics* 68: 1984–1999.
- [4]. Duda RO, Hart PO, Stork DG (2001) Pattern classification, 2nd ed. John Wiley & Sons.
- [5]. Taner MT, Koehler F, Sheriff RE (1979) Complex seismic trace analysis, *Geophysics* 44: 1041–1063.
- [6]. Balch AH (1971) Color sonagrams, A new dimension in seismic data interpretation, *Geophysics* 36, 1074–1098.
- [7]. Sonneland L (1983) Computer aided interpretation of seismic data, 53rd Annual International Meeting, SEG, Expanded Abstracts 546–549.
- [8]. Justice JH, Hawkins DJ, Wong DJ (1985) Multidimensional attribute analysis and pattern recognition for seismic interpretation, *Pattern Recognition* 18: 391–399.
- [9]. Barnes AE, Laughlin KJ (2002) Investigation of methods for unsupervised classification of seismic data, 72nd Annual International Meeting, SEG, Expanded Abstracts 2221–2224.
- [10]. Zhao T, Jayaram V, Roy A, Marfurt KJ (2015) A comparison of classification techniques for seismic facies recognition,": Interpretation 3, SAE29-SAE58.
- [11]. Chopra S, Marfurt KJ (2018) Seismic facies classification using some unsupervised machine learning methods," 2018 SEG Convention at Anaheim, California.
- [12]. Hadiloo S, Mirzaei S, Hashemi H, Beiranvand B (2018) Comparison Between Unsupervised and Supervise Fuzzy Clustering Method in Interactive Mode to Obtain the Best Result for Extract Subtle Patterns from Seismic Facies Maps, *Geopersia* 8:1, 27-34.
- [13]. MacQueen J (1967) Some methods for classification and analysis of multivariate observations, in L. M. Le Cam, and J. Neyman, eds., Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, University of California Press 281–297.
- [14]. Ferreira DJA, Lupinacci WM, de Andrade Neves I, Zambrini JPR, Ferrari AL, Gamboa LAP, Azul MO (2019) Unsupervised seismic facies classification applied to a presalt carbonate reservoir, Santos Basin, offshore Brazil, *AAPG Bulletin* 103:4, 997-1012.
- [15]. Coleou T, Poupon M, Azbel K (2003) Unsupervised seismic facies classification: A review and comparison of techniques and implementation, *The Leading Edge* 22, 942–953.
- [15]. Coleou T, Poupon M, Azbel K (2003) Unsupervised seismic facies classification: A review and comparison of techniques and implementation, *The Leading Edge* 22, 942–953.
- [16] Barnes A E (2015) Redundant and useless seismic attributes, *Geophysics* 72, P33-P38, 2007.
- [17]. Roden R, Smith T, Sacrey D (2015) Geologic Pattern Recognition from seismic attributes: principal component analysis and self-organizing maps, *Interpretation* 3, SAE59-SAE83.
- [18]. Gao D (2007) Application of three-dimensional seismic texture analysis with special reference to deep-marine facies discrimination and interpretation: An example from offshore Angola, West Africa, *AAPG Bulletin* 91, 1665–1683.
- [19]. Qian F, Yin M, Liu XY, Wang YJ, Lu C, Hu GM (2018) Unsupervised seismic facies analysis via deep convolutional autoencoders, *Geophysics* 83, 3: A39-A43.
- [20]. Bishop CM, Svensen M, Williams CKI (1998) The generative topographic mapping, *Neural Computation*, 10: 1, 215-234.
- [21]. Roy A (2013) Latent space classification of seismic facies, Ph.D. Dissertation, The University of Oklahoma, 212.
- [22]. Roy A, Romero-Pelaez AS, Kwiatkowski TJ, Marfurt K (2014) Generative topographic mapping for seismic facies estimation of a carbonate wash, Veracruz Basin, southern Mexico, *Interpretation* 2, SA31-SA47.
- [23]. Lubo D, Marfurt K, Jayaram V (2014) Statistical characterization and geological correlation of wells using automatic learning Gaussian mixture models, *Unconventional Resources Technology Conference, Extended Abstracts* 774–783.
- [24]. Hardisty R, Wallet B (2017) Unsupervised seismic facies from mixture models to highlight channel features, *SEG Technical Program Expanded Abstracts*, 2289-2293.
- [۲۵]. عدالت ع. و سیاهکوهی ح. ر.، "استفاده از رخساره‌های لرزه‌ای در توصیف یکی از مخازن نفتی ایران"، *مجله ژئوفیزیک ایران* ۱ (۱)، ۳۷-۴۹، ۱۳۸۶.
- [۲۶]. جوکار ع.، رحیمی م. و میرشکاری ف.، "دسته‌بندی و آنالیز رخساره‌های لرزه‌ای به روش‌های نظارتی و غیرنظارتی براساس آنالیز چند نشان‌گری"، *ماهنامه علمی-ترویجی اکتشاف و تولید نفت و گاز* ۵۰، ۶۷-۷۲،

۱۳۸۷. [۲۷]. هادیلو س.، سیاهکوهی ح. و عدالت ع.، "بهبود تحلیل رخساره‌های لِرزه‌ای با استفاده از نشان‌گر *WTM-MLA* و خوشه‌بندی‌های *SOM* و *K-mean*"، مجله فیزیک زمین و فضا، ۳۸ (۲)، ۴۵-۵۶، ۱۳۹۱.
- [۲۸]. ثابتی ح.، جواهریان ع. و نجار اعرابی ب.، "بررسی تغییر رخساره لِرزه‌ای براساس خوشه‌بندی سلسله‌مراتبی نشان‌گرهای لِرزه‌ای: بررسی موردی در یکی از میدان‌های نفتی ایران"، مجله ژئوفیزیک ایران ۸ (۴)، ۷۷-۹۶، ۱۳۹۳.
- [۲۹]. باقری م.، ریاحی م. ع.، هاشمی ح. و بختیاری م. ر.، "تحلیل رخساره‌های مخزن با استفاده از داده‌های لِرزه‌ای در فضای عدم تشابه"، ماهنامه علمی - ترویجی اکتشاف و تولید نفت و گاز، ۱۱۸، ۵۴-۵۹، ۱۳۹۳.
- [30]. Hartley H (1958) Maximum likelihood from incomplete data, *Biometrics* 14, 174-194.
- [31]. Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via the EM algorithm, *Journal of the Royal Statistical Society Series B (Methodological)* 39: 1: 1-38.
- [32]. Han M, Zhao Y, Li G, Reynolds AC (2011) Application of EM algorithms for seismic facies classification, *Computational Geosciences* 15: 421-429.
- [33]. Kung S, Mak M, Lin S (2004) *Biometric Authentication: A Machine Learning Approach*, Prentice Hall.
- [34]. Ambroise C, Dang M, Govaert G (1998) Convergence of an EM-type algorithm for spatial clustering, *Pattern recognition letters* 19: 919-927.
- [35]. Han M (2008) Application of the EM algorithms for facies classification and measurement error estimation, M.S. Thesis, University of Tulsa, Tulsa, Oklahoma, USA.
- [36]. Sfidari E, Kadkhodaie-Ilkhchi A, Rahimpour-Bbonab H, Soltani B (2014) A hybrid approach for litho-facies characterization in the framework of sequence stratigraphy: a case study from the South Pars gas field, the Persian Gulf basin, *Elsevier* 1: 121:87-102.
- [37]. Bezdek JC (1981) Cluster Validity. In *Pattern Recognition with Fuzzy Objective Function Algorithms*, 95-154, Boston, MA, Springer US.
- [38]. Hadiloo S, Hashemi H, Mirzaei S, Beiranvand B (2017) SeisART software: seismic facies analysis by contributing interpreter and computer, *Arabian Journal of Geosciences* 10, 23: 519.
- [۳۹]. هادیلو س.، "تحلیل رخساره‌های لِرزه‌ای سه‌بعدی با استفاده از روش تعاملی خوشه‌بندی و کلاسه‌بندی منطق فازی و *ANFIS*"، رساله دکتری، پژوهشکده علوم پایه کاربردی، جهاد دانشگاهی، ۱۳۹۷.



Unsupervised Seismic Data Classification Using Gaussian Mixture Models

Amir Ariyan Nezhad¹, Mohammad Radad^{2*} and Saeed Hadiloo³

1. Petroleum and Chemistry Engineering Department, Islamic Azad University, Electronic Campus, Tehran, Iran.

2. Faculty of Mining, Petroleum and Geophysics Engineering, Shahrood University of Technology, Iran.

3. Research Institute of Applied Sciences (ACECR), Tehran, Iran

mradad@shahroodut.ac.ir

DOI: 10.22078/pr.2020.3843.2755

Received: September/10/2019

Accepted: May/05/2020

Introduction

Description and interpretation of extracted seismic reflection parameters including geometry, continuity, amplitude, frequency and interval speed, are known as seismic facies analysis. Seismic facies analysis can be performed by pattern recognition techniques. By appropriate combination of seismic attributes, lateral changes in reservoir can be identified and then calibrated with well data. When the geological information is incomplete or nonexistent, seismic facies analysis is called unsupervised, and it is performed through unsupervised learning or clustering algorithms. Practically, unsupervised learning uses attributes as learning data and analyzed data. A comparison between different unsupervised methods for classification of seismic facies has been carried out by Barnes and Laughlin [1], Zhao et al. [2], Chopra and Marfurt [3] and Hadiloo et al [4]. Important unsupervised classification methods can include K-mean, Cross-plotting, Principal Component Analysis (PCA), Self-Organizing Maps (SOM), Generative Topographic Maps (GTM) and Gaussian Mixture Models (GMM) [2]. In this study, GMM is used for seismic facies analysis through unsupervised clustering in a set of 3D seismic data related to one of the hydrocarbon fields in south of Iran.

Materials and Methods

In a popular look out, clustering methods are classified in two categories: Classical (hard) and Fuzzy (soft)

clustering; In hard clustering, each data sample can only belong to one cluster; therefore, the sample does not have the probability of existence in two or more clusters. In other words, it is mentioned that the probability of lying each datum in special clusters is zero or one. However, in soft clustering, positioning a sample data set on more than one cluster is possible. GMM is a soft clustering technique. The approach of this method is to implement the expectation maximization (EM) algorithm for Gaussian mixture models. In essence, the EM algorithm is an iterative optimization method for parameter estimation. There are two steps at each iteration: 1) With the estimate of the Gaussian mixture model parameters from the previous iteration, the membership information (i.e. the probability that each datum is sampled from each Gaussian model) is estimated. This step is referred to as the expectation step, or simply the E-step. 2) Using the estimated membership information and the measurements, a maximum likelihood estimate of the Gaussian mixture model parameters is obtained in the second step. This step is referred to as the maximization step or simply the M-step. This two-step process is updated at each iteration until convergence is achieved.

Gaussian mixture model

In statistics, a Gaussian mixture model is a probability distribution that is a convex combination of some Gaussian probability distributions. Suppose that a random measurement d represents a sample from

a mixture of M given Gaussian models, then the probability density function (pdf) of d given the Gaussian mixture model is given by

$$P(d|II, \Theta) = \sum_{j=1}^M \pi_j P(d | \mu_j, C_j) \quad (1)$$

where $\Pi = (\pi_1, \pi_2, \dots, \pi_M)$ are the mixture proportions and the convex combination requires that $\pi_j \geq 0$ for each $j=1,2,\dots,M$ and $\sum_{j=1}^M \pi_j = 1$ [14]. $P(d | \mu_j, C_j)$ is a Gaussian distribution with mean μ_j and covariance matrix C_j . In Eq. 1, $\Theta = (\theta_1, \theta_2, \dots, \theta_M)$ and $\theta_j = (\mu_j, C_j)$. Let $\Phi = (II, \Theta)$ denote the Gaussian mixture model parameters, then Eq. 1 becomes:

$$P(d|\Phi) = \sum_{j=1}^M \pi_j P(d | \theta_j) \quad (2)$$

In the current research setting, it is assumed that seismic data of each geological facies follow a Gaussian probability distribution, therefore, the seismic data of the whole reservoir, which is composed of several facies, are samples of a Gaussian mixture model. To identify geological facies from seismic data, we need to find the Gaussian models and the model from which each datum is sampled [5]. This method is performed by using EM algorithm where the details of theory is described in Han et al [5] and Kung et al [6] researches [5,6].

Applying GMM to seismic data

By applying GMM on seismic attributes, classification of geological different features can be achieved. In this section, 3D seismic information is clustered in one of hydrocarbon fields in south of Iran, and the performance of this method for seismic facies analysis is investigated. Two continuous horizons from 3D seismic data called K1 and K2 are selected, and seismic facies analysis is applied on each horizon. In facies analysis, using GMM, seismic facies distribution is achieved and also membership probability of each member to its own facies is defined by a digit between 0 and 1, where by using probability distribution, the performance and accuracy of clustering can be checked. In order to extract seismic information on desired horizons, result of picking on each horizon is used to extract seismic information and seismic attributes. In order to perform a suitable clustering, an optimized combination of seismic attributes should be selected and be used. In this study, different attributes are used including

instantaneous amplitude, instantaneous frequency, instantaneous phase, dominant frequency, which have basic information from seismic data, and texture attributes that represent probable textural changes in horizons, and also single frequency attributes of 20, 30 and 40 Hertz obtained from spectral decomposition of seismic data that have frequency change information and relative acoustic impedance attribute which has information related to acoustic impedance. In order to analyze seismic facies on each horizon, mentioned attributes are generated and after extracting information on desired horizon, GMM method is applied to cluster and generate seismic facies map.

Facies analysis using GMM

In order to cluster seismic information using GMM, number of clusters (or number of seismic facies) for analyzing seismic facies should be defined. This can be done by using previous information of the area, or the appropriate number of clusters for the desired information can be obtained by using mathematical methods of defining optimized number of clusters [7,8]. In this study, based on the research of Hadiloo et al [8] on this zone, the number of five clusters is selected for facies analysis of this data. Facies analysis maps from clustering seismic information in two desired horizons using GMM method by considering five clusters and a cluster membership probability map are provided for each horizon analysis.

Results and Discussion

The results of horizon K1 analysis are shown in Figures 1 and 2, and those of horizon K2 analysis are presented in Figures 3 and 4. In the cluster membership probability maps, whatever the probability of presentation is close to 1, it represents that the data are more probable member of desired cluster, and the data are near center of gaussian distribution. This map can be used appropriately to determine the accuracy and capability of the clustering method. When all members have high probability and close to one (dark red), it can be mentioned that the attributes and the number of clusters are appropriately selected. For each of the two horizons, it can be qualitatively seen that probability distributions of membership are approximately close to one, and this can show that GMM clustering could perform appropriate classification of seismic data.

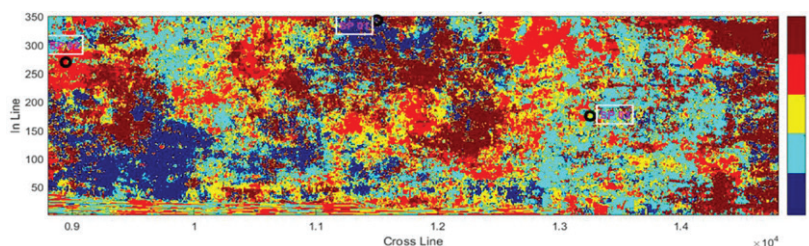


Fig. 1 Facies classification of the horizon K1, clustered with GMM method by 5 clusters. Three wells location are remarked by black circles.

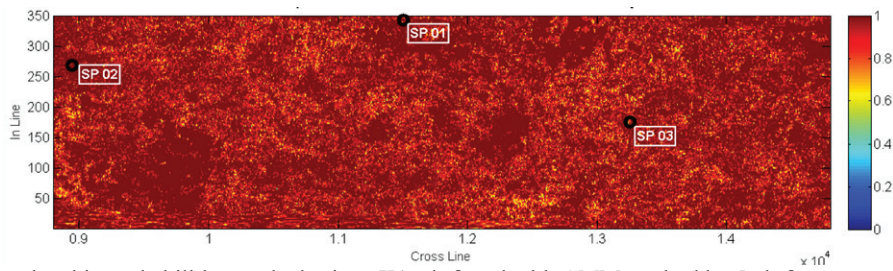


Fig. 2 Cluster membership probabilities on the horizon K1, clustered with GMM method by 5 clusters.

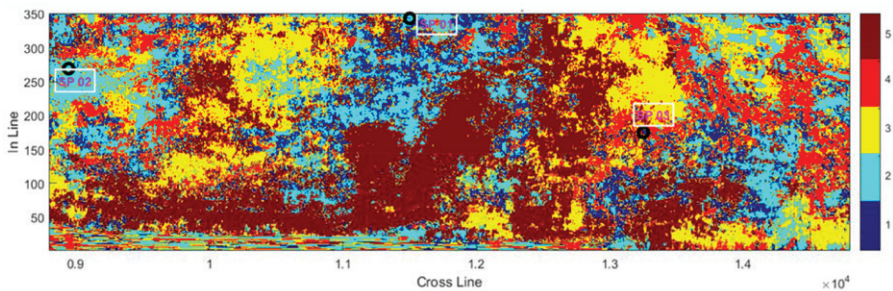


Fig. 3 Facies classification of the horizon K2, clustered with GMM method by 5 clusters. Three wells location are remarked by black circles.

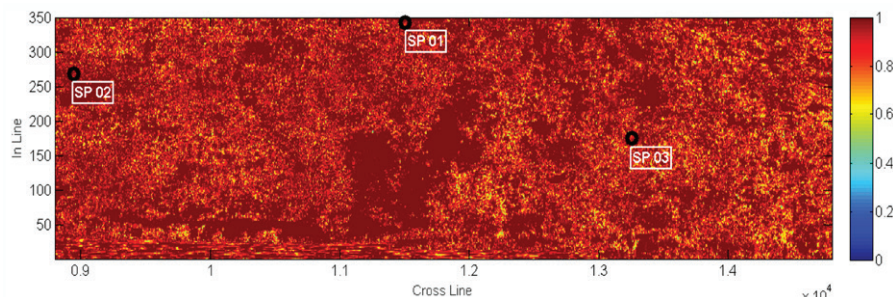


Fig.4. Cluster membership probabilities on the horizon K2, clustered with GMM method by 5 clusters.

Conclusions

In this study, GMM method application was used for unsupervised seismic facies analysis. GMM clustering method is able to identify regularity in the data and distribution of seismic facies without using well information. In GMM method, gaussian functions are used for estimating and covering seismic information. Finally, according to this study, it is found out that GMM method can generate seismic facies distribution with an acceptable correlation between neighbor horizons for 3D seismic information in two different horizons, and also present that the method allows interpreter to comment qualitatively about quality of applied clustering by providing a cluster membership probability map.

Nomenclatures

GMM: Gaussian Mixture Models
 GTM: Generative Topographic Maps
 PCA: Principal Component Analysis

References

1. Barnes AE, Laughlin KJ (2002) Investigation of

methods for unsupervised classification of seismic data, 72nd Annual International Meeting, SEG, Expanded Abstracts, 2221–2224.

2. Zhao T, Jayaram V, Roy A, Marfurt KJ (2015) A comparison of classification techniques for seismic facies recognition, Interpretation, 3, SAE29-SAE58.
3. Chopra S, Marfurt KJ (2018) Seismic facies classification using some unsupervised machine learning methods, 2018 SEG Convention at Anaheim, California.
4. Hadiloo S, Mirzaei S, Hashemi H, Beiranvand B (2018) Comparison Between Unsupervised and Supervise Fuzzy Clustering Method in Interactive Mode to Obtain the Best Result for Extract Subtle Patterns from Seismic Facies Maps, Geopersia, 8(1), pp.27-34.
5. Han M, Zhao Y, Li G, Reynolds AC (2011) Application of EM algorithms for seismic facies classification, Comput. Geosci., 15, 421-429.
6. Kung S, Mak M, Lin S (2004) Biometric Authentication: A Machine Learning Approach, Prentice Hall.

7. Bezdek JC (1981a) Cluster Validity. In Pattern Recognition with Fuzzy Objective Function Algorithms (pp. 95–154). Boston, MA, Springer US.
8. Hadiloo S, Hashemi H, Mirzaei S., Beiranvand B (2017) SeisART software: seismic facies analysis by contributing interpreter and computer, Arabian Journal of Geosciences, 10(23), p.519.