

ساخت نگار لیتولوژی با رویکرد خوشه‌بندی - تخمین در چاه‌های یک میدان نفتی جنوب ایران

ساره صدیق^{۱*}، مهرانوش علی‌پور شهسواری^۱، حسین معماریان^۱، بهزاد تخم‌چی^۲

^۱دانشکده فنی، دانشگاه تهران

^۲دانشگاه صنعتی شاهرود

*sarehsadigh@yahoo.com

تاریخ دریافت: خرداد ۹۲، تاریخ پذیرش: شهریور ۹۲

چکیده

تعیین لیتولوژی مخزن از مهم‌ترین بررسی‌های مخزنی است که جهت تطابق چاه‌ها و تشخیص قسمت‌های تولیدی مخزن به کار می‌رود. بهترین روش در تعیین لیتولوژی استفاده از اطلاعات مغزه و خرده‌های حفاری است. اما در بسیاری از چاه‌ها این اطلاعات به صورت کامل و پیوسته موجود نیست. به طور معمول در موارد نبود مغزه نسبت به تخمین لیتولوژی از نگارهای پتروفیزیکی استفاده می‌شود. هدف این مقاله ارائه‌ی روش دقیق‌تری جهت تخمین لیتولوژی از داده‌های پتروفیزیکی با رویکرد خوشه‌بندی - تخمین است. این روش بر اساس داده‌های یک چاه اکتشافی از یکی از میداین نفتی جنوب ایران که دارای نتایج آنالیز مغزه بوده توسعه داده شده و نسبت به بهینه‌سازی پارامترهای مدل اقدام شده است. سپس این مدل بر روی چاه‌های فاقد مغزه تعمیم داده شده است. خوشه‌بندی به عنوان عاملی برای تفکیک داده‌های چاه به جوامع همگن لیتولوژیکی مورد استفاده قرار می‌گیرد، سپس تخمین درصد کانی‌ها در هر کدام از این جوامع غالب لیتولوژیکی صورت گرفته است و به ترتیب ضرایب همبستگی ۹۲/۹۳٪ و ۷۴/۹۹٪ بین داده‌های واقعی و تخمینی دولومیت و کلسیت در یکی از چاه‌ها به دست آمده است. نتایج معرف دقت مناسب و قابلیت تعمیم قابل توجه رویکرد است.

کلمات کلیدی: مخزن آسماری، خوشه‌بندی، لیتولوژی، تخمین گر MLP، خوشه‌بندی میانگین k داده، خوشه‌بندی گوستاوسون کسل، ایران.

۱. مقدمه

مهم‌ترین روش‌های تعیین لیتولوژی، استفاده از چاه‌نمودارهای پتروفیزیکی، مطالعه مغزه‌ها و استفاده از نمودارهای استاندارد تعیین لیتولوژی است. استفاده از مغزه‌ها روشی مستقیم در شناسایی لیتولوژی به حساب می‌آید، اما تهیه مغزه بسیار پرهزینه است و همچنین به علت تهیه ناپیوسته‌ی آنها در طول یک چاه، معمولاً شرح کاملی از سازند را در بر نخواهند داشت. این در حالی است که چاه‌نمودارها می‌توانند یک برآورد پیوسته از یک سازند را در طول چاه مشخص نمایند. محققانی همچون دلفینر^۱، کلاویر^۲ و روست^۳ نشان دادند که پاسخ چاه‌نمودارها ایده‌ی مناسبی جهت تعیین لیتولوژی است [۱ و ۲].

پیش از پیدایش کامپیوترهای امروزی متداول‌ترین روش تعیین لیتولوژی استفاده از نمودارهای استاندارد بود. پرکاربردترین این نمودارها توسط شرکت شلومبرژه^۴ ارائه شده است. بالغ بر ۱۰ نمودار دویعدی برای تعیین لیتولوژی وجود دارد که هر کدام بر اساس ویژگی‌های پتروفیزیکی گوناگونی نظیر چگالی حجمی، اثر فتوالکتریک و غیره قادر به تعیین کانی‌های مخزنی هستند [۳]. اما این نمودارها تنها قادر به استفاده از دو ویژگی پتروفیزیکی جهت تعیین لیتولوژی بودند و در تعیین ساختارهای لیتولوژیکی ساده کارایی داشتند. در مجموع در این نمودارها، نظر کارشناسان اهمیت بیشتری داشت. با پیدایش نسل جدید کامپیوترها و ارتقای برنامه‌های پردازشی، قدرت پردازش داده‌ها چند برابر شد و طی سه دهه‌ی اخیر استفاده از روش‌های هوشمند جایگزینی مناسب برای تعیین لیتولوژی شد. مهم‌ترین مزیت استفاده از این تکنیک‌ها، قدرت افزایش فضای ویژگی و در نتیجه افزایش دقت پردازش است. چاه‌نمودارهای اثر فتوالکتریک، چگالی، اشعه‌ی گاما و تخلخل نوترون مهم‌ترین چاه‌نمودارها در تخمین لیتولوژی هستند که به عنوان ورودی در خوشه‌بندی و تخمین‌گرها مورد استفاده قرار می‌گیرند. برخی محققین نظیر کودی^۵ و اجبای^۶ (سال ۲۰۱۲) اظهار داشتند که می‌توان به کمک چاه نمودار اشعه‌ی گاما، لیتولوژی‌های غالب مانند ماسه و شیل را از یکدیگر تفکیک نمود و زون‌های تراوا و متخلخل را از زون‌های ناتراوا و کم تخلخل جدا کرد [۴]. محققان دیگری نظیر آکینوکان^۷ و همکاران (سال ۲۰۰۹) با استفاده از چاه نمودارهای چگالی، اشعه‌ی گاما و تخلخل نوترون و به‌کارگیری روش‌های شبکه‌ی عصبی به تخمین لیتولوژی پرداخته‌اند و موفق به تفکیک لیتولوژی‌های غالب نظیر ماسه‌سنگ و شیل شدند، علاوه بر آن با استفاده از این تکنیک‌ها و به‌کارگیری نمودارهای مقاومت ویژه‌ی سازند به محتوای سیال نیز پی بردند [۵].

با توجه به اهمیت تعیین لیتولوژی در ارزیابی ذخایر هیدروکربنی و سنجش کیفیت مخزن، هدف این مقاله بهینه‌سازی روش‌های خوشه‌بندی در جهت تعیین جوامع غالب لیتولوژیکی و ارائه‌ی رویکرد خوشه‌بندی-تخمین جهت تخمین لیتولوژی است. به این منظور در ابتدا به معرفی روش تحقیق، که شامل الگوریتم خوشه‌بندی و تخمین است پرداخته شده است، در مرحله‌ی بعد داده‌های مورد استفاده و ارزیابی اولیه‌ی آنها ارائه شده است. در خاتمه نیز نتایج اعمال خوشه‌بندی و تخمین‌گر بر روی داده‌ها نشان داده شده است.

¹ Delfiner

² Clavier

³ Rust

⁴ Schlumberger

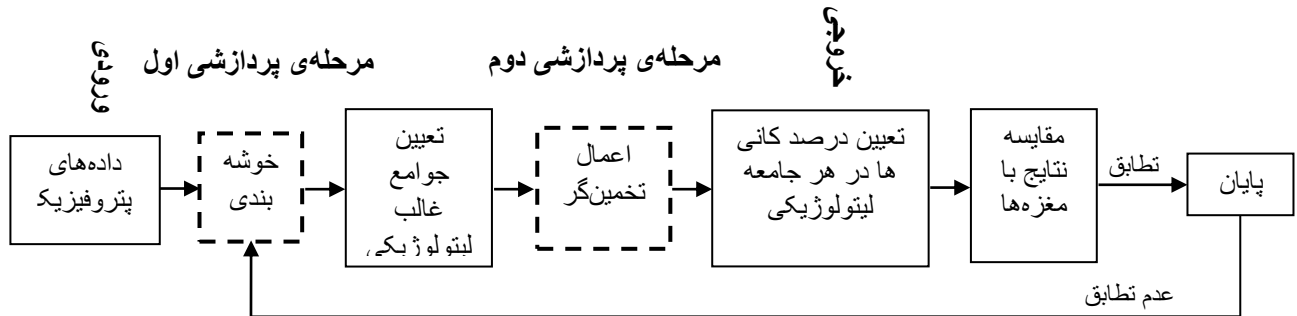
⁵ Cuddy

⁶ Egbai

⁷ Akinyokun

۲. روش تحقیق

فرایند پژوهش حاضر شامل سه بخش ورود داده‌های پتروفیزیکی، بخش پردازشی و دریافت خروجی‌ها است (شکل ۱).



شکل ۱: نمایش گرافیکی از مراحل انجام تحقیق

بخش پردازشی خود از دو زیربخش اساسی تشکیل شده است. قسمت اول مربوط به استفاده از خوشه‌بندی جهت تعیین جوامع غالب لیتولوژیکی و همگن کردن فضای داده است. روش خوشه‌بندی استفاده شده در این مقاله الگوریتم گوستاوسون کسل^۱ و تخمین‌گر مورد استفاده شبکه‌ی پرسپترون چند لایه (MLP) است.

۲-۱۱ الگوریتم گوستاوسون کسل

الگوریتم گوستاوسون کسل یکی از روش‌های خوشه‌بندی فازی به‌شمار می‌آید که در آن برخلاف روش‌های خوشه‌بندی معمول نظیر الگوریتم میانگین k داده^۲، برای سنجش میزان شباهت بین نمونه‌ها، از معیار فاصله‌ی ماهالانوبیس^۳ استفاده می‌شود. مزیت این معیار، تفکیک خوشه‌ها با توجه به ساختار داده‌ها است و این عمل با محاسبه‌ی ماتریس القاگر داده‌ها در خوشه‌ها صورت می‌گیرد همین امر سبب می‌شود تا الگوریتم قادر به تشخیص خوشه‌های با ساختارهای خطی و بیضوی نیز باشد [۶]. مراحل پردازشی الگوریتم به شرح زیر است [۷]:

الف- قبل از وارد کردن داده‌ها در الگوریتم نیاز است تا تمامی آنها در دامنه‌ی عددی مشابهی قرار گرفته تا بتوان برای پردازش‌های بعدی از آنها استفاده نمود. برای نرمال کردن داده‌ها از معادله‌ی (۱) استفاده شده است:

$$\bar{x}_i = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}}$$

که در آن x_i داده‌ی نرمال نشده، x_{\max} و x_{\min} به ترتیب کمترین و بیشترین داده‌ی نرمال نشده و \bar{x} داده‌ی نرمال شده است.

ب- تعیین تعداد خوشه‌های c ، و مقدار دهی پارامتر وزنی m که معمولاً برابر با دو در نظر گرفته می‌شود و میزان فازی بودن داده‌ها را نشان می‌دهد.

ج- تعیین مقدار اولیه ماتریس جداسازی $U_0 = [u_{ik}]$ به صورت تصادفی.

د- محاسبه‌ی مراکز خوشه‌ها V_i مطابق معادله‌ی (۲):

$$V_i = \frac{\sum_{k=1}^n (u_{ik})^m x_k}{\sum_{k=1}^n (u_{ik})^m}, \quad i \in [1, c] \quad \text{and} \quad k \in [1, n]$$

^۱ Gustafson Kessel

^۲ برای کسب اطلاعات بیشتر در ارتباط با عملکرد این الگوریتم می‌توان به Webb A., 2002 Statistical Pattern Recognition, مراجعه نمود.

^۳ Mahalanobis

که در آن n تعداد دفعات محاسبه‌ی ماتریس جداسازی و محاسبه‌ی مرکز خوشه است.
 ه- محاسبه‌ی ماتریس کواریانس F_i که نشان‌دهنده‌ی چگونگی پراکندگی داده‌ها در یک خوشه است و نحوه‌ی محاسبه‌ی آن برای هر خوشه مطابق با معادله‌ی (۳) است:

$$F_i = \frac{\sum_{k=1}^n (u_{ik})^m (x_k - v_i)(x_k - v_i)^T}{\sum_{k=1}^n (u_{ik})^m}, \quad i \in [1, c]$$

و- محاسبه‌ی ماتریس الفاکر A_i برای هر خوشه مطابق معادله‌ی (۴):

$$A_i = \sqrt[n]{\rho_i \det(F_i)} F_i^{-1}, \quad i \in [1, c]$$

که در آن h بعد فضای داده و ρ_i برابر با ۱ در نظر گرفته می‌شود.

ز- محاسبه‌ی فاصله‌ی ماهالانویس برای هر نمونه x_k از هر مرکز خوشه v_i مطابق با معادله توسط معادله‌ی (۵):

$$d_{ik}^2 = (x_k - v_i)^T A_i (x_k - v_i)$$

ح- جدید کردن ماتریس جداسازی $U = [u_{ik}]$ برای هر نمونه و تعیین درجه‌ی عضویت نمونه‌ها مطابق با معادله‌ی (۶):

$$u_{ik} = \frac{1}{\sum_{j=1}^c \left(\frac{d_{ik}}{d_{ij}} \right)^{\frac{2}{m-1}}}, \quad i \in [1, c] \text{ and } k \in [1, n]$$

پس از تعیین جوامع غالب لیتولوژیکی به کمک روش‌های خوشه‌بندی، بخش دوم پردازش استفاده از تخمین‌گر MLP جهت تخمین درصد کانی‌هاست.

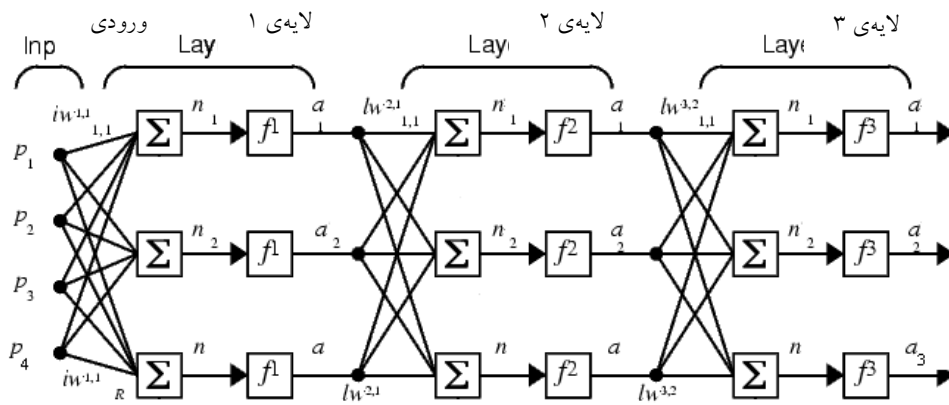
۲-۲ تخمین‌گر MLP

یک شبکه‌ی عصبی مصنوعی (ANN)^۱ ابزاری محاسباتی موازی برای پردازش اطلاعات است که از سیستم عصبی زیستی الهام گرفته شده و مانند مغز به پردازش اطلاعات می‌پردازد. عنصر کلیدی این ایده، ساختار جدید سیستم پردازش اطلاعات است. یک شبکه‌ی عصبی قادر به انجام پردازش‌هایی نظیر شناسایی الگوها و دسته‌بندی اطلاعات است [۸ و ۹]. در این مقاله هدف تعیین درصد کانی‌های مخزنی در بخش‌های فاقد مغزه است، لذا ابتدا به کمک داده‌های مغزه به شبکه‌ی عصبی آموزش داده می‌شود و سپس از آن خواسته می‌شود تا در بخش‌های دیگر درصد کانی‌ها را تخمین بزند. عناصر در شبکه‌های عصبی به طرق مختلفی به یکدیگر اتصال پیدا می‌کنند و ساختارهای گوناگونی را پدید می‌آورند. یکی از این ساختارها، شبکه‌ی پرسپترون چند لایه (MLP)^۲ است.

^۱ Artificial Neural Network

^۲ Multilayer Perceptron

این شبکه‌ها قادرند با انتخاب مناسب تعداد لایه‌ها و سلول‌های عصبی، یک نگاشت غیرخطی را با دقت دلخواه انجام دهند. شبکه‌های پرسپترون تک لایه از این مشکل اساسی برخوردارند که تنها توانایی حل آن دسته از مسائل طبقه‌بندی را داشتند، که به طور خطی از هم متمایز می‌شدند. برای حل این مشکل شبکه‌های پرسپترون چند لایه مطرح شد. وجود لایه‌ی میانی، قابلیت شبکه‌های پرسپترون چند لایه را در تخمین یک نگاشت افزایش می‌دهد. افزایش تعداد لایه میانی شبکه را قادر می‌سازد تا اطلاعات آماری کامل‌تری در مورد مسئله کسب کند [۱۰]. به عنوان مثال در شکل ۲ یک شبکه‌ی عصبی با چهار ورودی، سه لایه و سه خروجی است. ورودی‌های p_1 تا p_4 وارد لایه ورودی می‌شوند و پس از ضرب هر ورودی در مقدار وزن خود و محاسبه‌ی ورودی کل، وارد تابع هر نرون می‌شوند. هر نرون یک خروجی تولید می‌کند که این خروجی به عنوان ورودی، وارد تمام نرون‌های لایه بعدی می‌شود. بدین ترتیب، خروجی‌های لایه اول، بردار ورودی لایه دوم را تشکیل می‌دهند و به همین ترتیب بردار خروجی لایه دوم، ورودی‌های لایه سوم را می‌سازند و خروجی‌های لایه سوم، پاسخ واقعی شبکه را تشکیل می‌دهند. در نهایت خروجی‌های a_1 تا a_3 به عنوان خروجی‌های شبکه حاصل می‌شوند [۱۱].



شکل ۲: مثالی از شبکه‌ی عصبی پرسپترون چند لایه [۱۱]

مراحل کلی در الگوریتم MLP به شرح زیر است [۱۰ و ۱۲]:

الف- تعیین متغیرهای ورودی و خروجی مطلوب.

ب- نرمال‌سازی داده‌ها در دامنه‌ی دلخواه که این مرحله مطابق با معادله‌ی (۱) در قسمت خوشه‌بندی صورت گرفته است.

ج- انتخاب تصادفی ۷۰٪ و ۳۰٪ داده‌ها به ترتیب به عنوان داده‌های آموزش و داده‌های آزمون.

د- تعیین تابع برای محاسبه‌ی خروجی. توابع برانگیزش متعددی برای محاسبه‌ی خروجی وجود دارد، که در این مقاله از تابع برانگیزش خطی مطابق معادله‌ی (۷) استفاده شده است که در آن خروجی به‌دست آمده برابر با ورودی آن است.

$$a = f(n) = \text{purelin}(n) = n$$

ه- شبکه مقادیر بردارهای وزن خود را در حین ارائه‌ی دسته داده‌های آموزشی و با استفاده از یک قانون یادگیری، طوری تنظیم می‌کند که با کمترین خطای ممکن رابطه بین خروجی‌ها و ورودی‌های الگوهای آموزشی را تشخیص دهد. این فرآیند را مرحله یادگیری یا آموزش شبکه می‌نامند.

و- با توجه به مشخص بودن خروجی، آموزش با نظارت بر روی داده‌ها صورت گرفته و از خروجی به عنوان بردار آموزش‌دهنده استفاده می‌شود. بعد از اعمال ورودی x_i به شبکه، خروجی y_i به عنوان تقریبی از مقدار حقیقی t_i ساخته

می‌شود و از خطای $(y_i - t_i)$ به عنوان عاملی برای اصلاح وزن‌ها استفاده می‌شود. خطای کلی شبکه از معادله‌ی (۸) به دست می‌آید:

$$E = \frac{1}{2} \sum_{p=1}^P \sum_{N=1}^N (y_i - t_i)^2$$

که در آن E خطای کلی شبکه، P تعداد الگوهای آموزشی، t_i خروجی واقعی، y_i خروجی شبکه در گره‌ی i و N تعداد گره‌های خروجی است. بر اساس خطای محاسبه شده‌ی وزن‌های شبکه آنقدر تغییر داده می‌شوند تا اختلاف خروجی شبکه و خروجی مطلوب به حداقل مورد نظر برسد.

ز- آموزش هنگامی متوقف می‌شود که مقادیر خروجی محاسبه شده بهترین تقریب را با مقادیر مطلوب داشته باشند. به عبارت دیگر، اگر مجموع خطاهای کلیه الگوهای یادگیری به یک مقدار کوچک از پیش تعیین شده همگرا شود، فاز یادگیری خاتمه می‌یابد. در غیر این صورت، انتشار پسرو انجام شده و خطای محاسبه شده به صورت پسرو در شبکه منتشر می‌شوند و وزن‌های ارتباطی را تصحیح می‌کند.

۳. داده‌ها

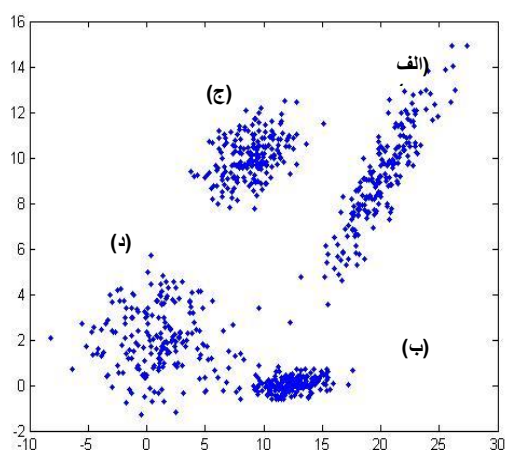
در این مقاله، به منظور بررسی میزان کارایی روش‌های خوشه‌بندی از دو سری داده‌های مصنوعی و واقعی استفاده شد. در ابتدا الگوریتم‌ها بر روی داده‌های مصنوعی اعمال و میزان کارایی و صحت پاسخ‌ها بررسی شد و با توجه به نتایج به دست آمده، عملیات خوشه‌بندی بر روی داده‌های واقعی مورد آزمایش قرار گرفت.

۳-۱- داده‌های مصنوعی

یکی از راه‌های کنترل نحوه‌ی عملکرد یک الگوریتم استفاده از داده‌های مصنوعی است. یک مجموعه داده‌ی مصنوعی همان‌طور که از اسم آن پیداست ساخته دست انسان است و اطلاع کامل در ارتباط با جنس داده، نوع توزیع آن، چگونگی پراکندگی و غیره در دست است. استفاده از داده‌های مصنوعی در واقع راهی برای ارزیابی یک کد و الگوریتم محاسباتی است [۱۳].

در این مقاله از یک مجموعه داده‌ی مصنوعی که داده‌های آن از چهار دسته با ساختارهای کروی، بیضوی و خطی تشکیل شده‌اند استفاده شده است. این مجموعه داده در نرم افزار متلب^۱ آماده سازی شده و با انتخاب یک ماتریس میانگین با درایه‌های 2×2 به عنوان مرکز هر خوشه و ماتریس کواریانس با درایه‌های 2×2 به عنوان عاملی برای تعیین نحوه‌ی پراکندگی داده‌ها و شکل هر خوشه برای هر یک از چهار خوشه، این مجموعه داده‌ی مصنوعی ساخته شد (شکل ۳).

¹ Matlab



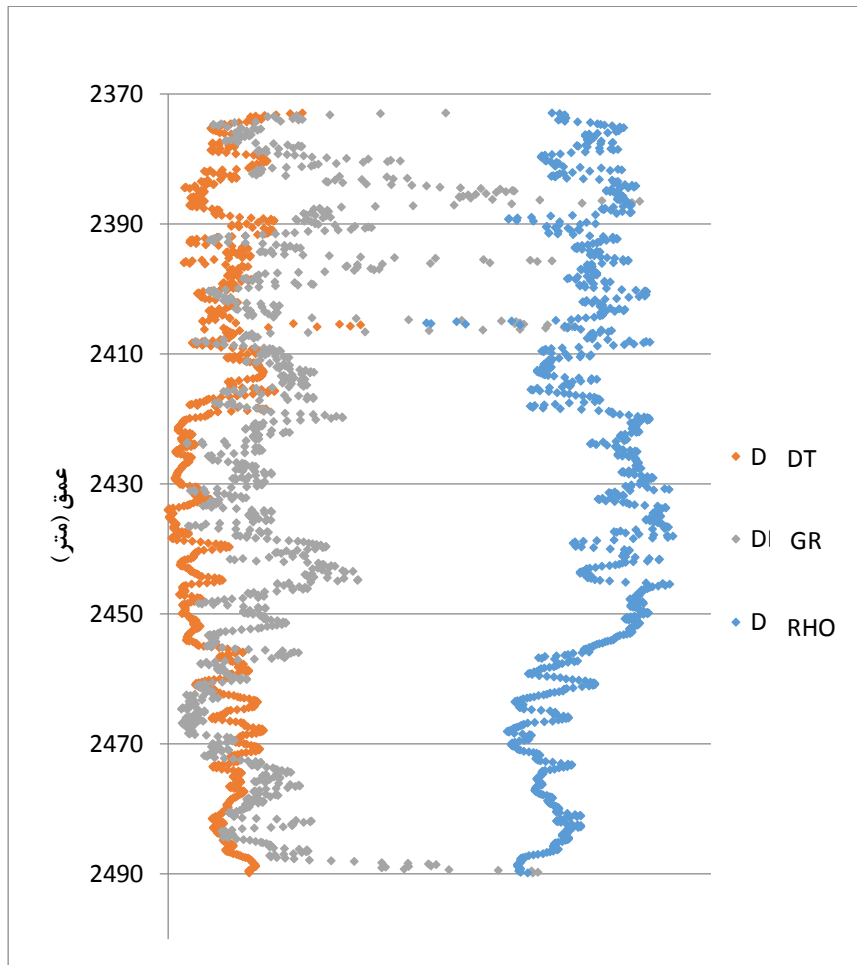
شکل ۳: داده‌های مصنوعی با ساختار (الف) - خطی، (ب) - بیضوی، (ج) و (د) - توده‌ای

۲-۳ داده‌های واقعی

داده‌های واقعی که در این مقاله از آن استفاده شد، داده‌های مربوط به اطلاعات لاگ‌های پتروفیزیکی مخزن آسماری در چاه شماره ۷ واقع در یکی از میداین نفتی جنوب ایران است (شکل ۴). نکته‌ی قابل اهمیت در انتخاب ویژگی‌ها، بررسی میزان اهمیت و تاثیر آنها در به دست آوردن پاسخ‌های صحیح و نزدیک به واقعیت است. به همین خاطر ابتدا می‌بایست از حجم وسیع اطلاعات موجود به انتخاب ویژگی^۲ پرداخت و پارامترهای تاثیر گذار بر تعیین لیتولوژی را از بین آنها انتخاب نمود.

¹ Features

² Feature Selection



شکل ۴: نمایی از چاه‌نمودارهای استفاده شده در تخمین لیتولوژی

همان طور که در شکل ۴ مشاهده می‌شود مهمترین ویژگی‌ها که در این مقاله برای تعیین لیتولوژی و درصد کانی‌های مخزنی به کار گرفته شده‌اند عبارت از نمودارهای چگالی حجمی (RHOB)، اشعه‌ی گاما (GR) و سرعت صوت (DT) است. همان‌طور که ذکر شد در این مقاله تنها نتایج بررسی بر روی چاه شماره‌ی ۷ از یک میدان نفتی آورده شده است و نمودارهای موجود در سایر چاه‌ها و محدوده‌ای عمقی آنها در جدول ۱ آورده شده است.

جدول ۱: مشخصات چاه‌های مورد استفاده در تخمین لیتولوژی

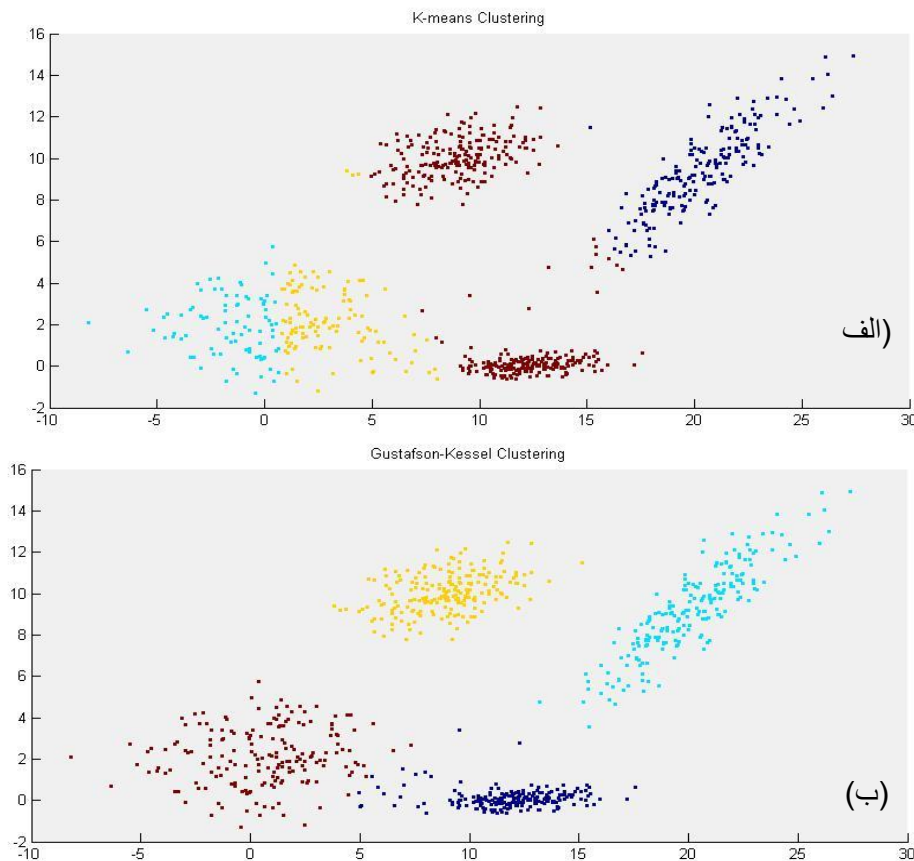
چاه‌نمودار RHOB	چاه‌نمودار PEF	چاه‌نمودار GR	چاه‌نمودار DT	عمق (m)	چاه شماره‌ی
✓	-	✓	✓	۲۵۱۵/۵ - ۲۳۸۹/۶	۱
✓	-	✓	✓	۲۵۲۴ - ۲۴۰۳	۲
✓	-	✓	✓	۲۵۲۷ - ۲۴۰۷	۳
✓	-	✓	✓	۲۵۱۲/۵ - ۲۳۸۶/۱	۴
✓	-	✓	✓	۲۵۲۳/۹ - ۲۳۹۸/۲	۵
✓	✓	✓	✓	۲۴۸۹/۸ - ۲۳۷۳	۷

۴. نتایج خوشه‌بندی

برای کاهش خطا در تخمین درصد کانی‌ها، ابتدا لازم بود تا با اعمال خوشه‌بندی، جوامع غالب لیتولوژیکی تعیین شود و سپس در هر یک از آنها تخمین‌ها صورت گیرد. الگوریتم‌های خوشه‌بندی بر روی دو گروه داده‌های مصنوعی و واقعی اعمال شده و نتایج آن در ادامه آمده است.

۴-۱ نتایج خوشه‌بندی داده‌های مصنوعی

داده‌های مصنوعی در چهار خوشه‌ی جدا از هم ساخته شده‌اند. همان‌طور که در شکل ۵-الف پیداست، الگوریتم میانگین k ، داده‌ها را به درستی تفکیک نکرده چرا که با توجه به تعریفی که در معیار فاصله‌ی آن وجود دارد، این الگوریتم تمایل به تشخیص خوشه‌های کروی و به عبارتی توده‌ای دارد، خواه این خوشه‌ها وجود خارجی داشته و یا نداشته باشند. اما مطابق شکل ۵-ب، الگوریتم گوستاوسون-کسل به خوبی خوشه‌ها را تشخیص داده و تفکیک بهینه‌تری نسبت به روش میانگین k داده انجام داده است.

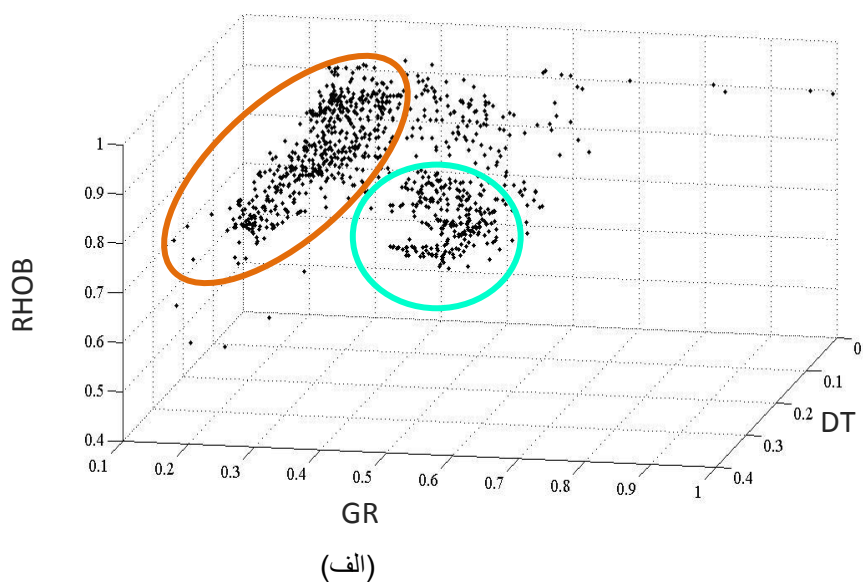


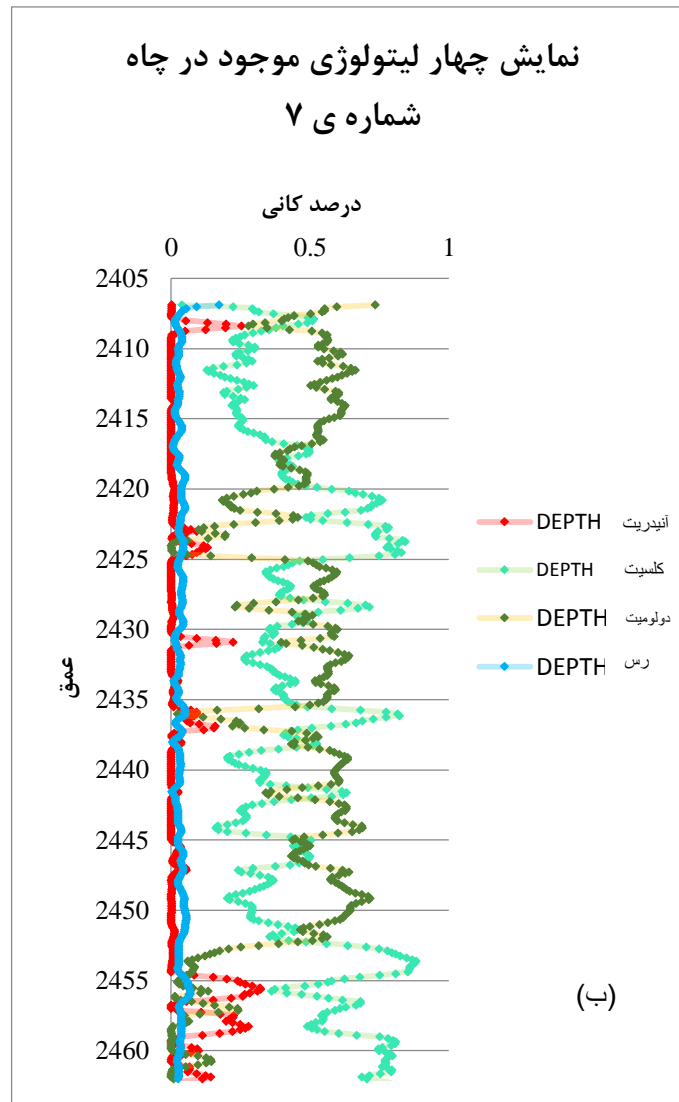
شکل ۵: خوشه‌بندی داده‌های مصنوعی، الف- به روش k میانگین‌داده، ب- به روش گوستاوسون-کسل

۴-۲ نتایج خوشه‌بندی داده‌های واقعی

همان‌طور که قبلاً اشاره شد، هدف از انجام خوشه‌بندی تعیین لیتولوژی‌های غالب در راستای چاه انتخابی‌ست، با توجه به داده‌های مغزه، در کل چاه انتخاب شده، ۴ نوع گوناگون از لیتولوژی وجود دارد، اما با انجام بررسی آماری بر روی داده‌های مغزه مشاهده شد که تنها دو نوع لیتولوژی در راستای این چاه غالب هستند. لذا مطابق شکل ۶-الف تعداد

اولیهی خوشه‌ها برابر ۲ انتخاب شد. این دو خوشه‌ی غالب لیتولوژیکی عبارتند از کلسیت و دولومیت. شکل ۶-ب نشان دهنده‌ی وجود چهار لیتولوژی در توالی مورد بررسی و غالب بودن دو لیتولوژی در راستای چاه شماره‌ی ۷ است.





شکل ۶- (الف) تفکیک دو لیتولوژی غالب کلسیت و دولومیت در نمودار سه‌بعدی DT-GR-RHOB. (ب) نشان‌دهنده‌ی چهار لیتولوژی در منطقه و غالب بودن دو لیتولوژی کلسیت و دولومیت در راستای چاه شماره‌ی ۷

۳-۴ شاخص‌های اعتبارسنجی^۱

از آنجا که خوشه‌بندی یک فرایند نظارت نشده است، لذا انتخاب الگوریتم امری بسیار مهم تلقی می‌شود و از مواردی که در خوشه‌بندی می‌بایست به آن توجه نمود، میزان عملکرد یک الگوریتم و انتخاب تعداد بهینه‌ی خوشه‌هاست، به همین منظور می‌توان از شاخص‌های اعتبارسنجی استفاده نمود [۱۴]. در این مقاله نیز برای اطمینان از انتخاب پارامترهای الگوریتم خوشه‌بندی از چهار شاخص اعتبارسنجی سیلهوته^۲، کالینسکی-هاراباسز^۳، کرزانوسکی-لای^۴ و دیویس-بولدین^۵ استفاده شده است. به‌طور کلی این شاخص‌ها بر اساس فواصل درون خوشه‌ای و برون خوشه‌ای تعریف می‌شود. معیار سیلهوته براساس میانگین فاصله هر یک از نمونه‌های یک خوشه با تمام نمونه‌های موجود در همان خوشه و

¹ Validity Index

² Silhouette

³ Calinski&Harabasz

⁴ Krzanowski&Lai

⁵ Davies & Bouldin

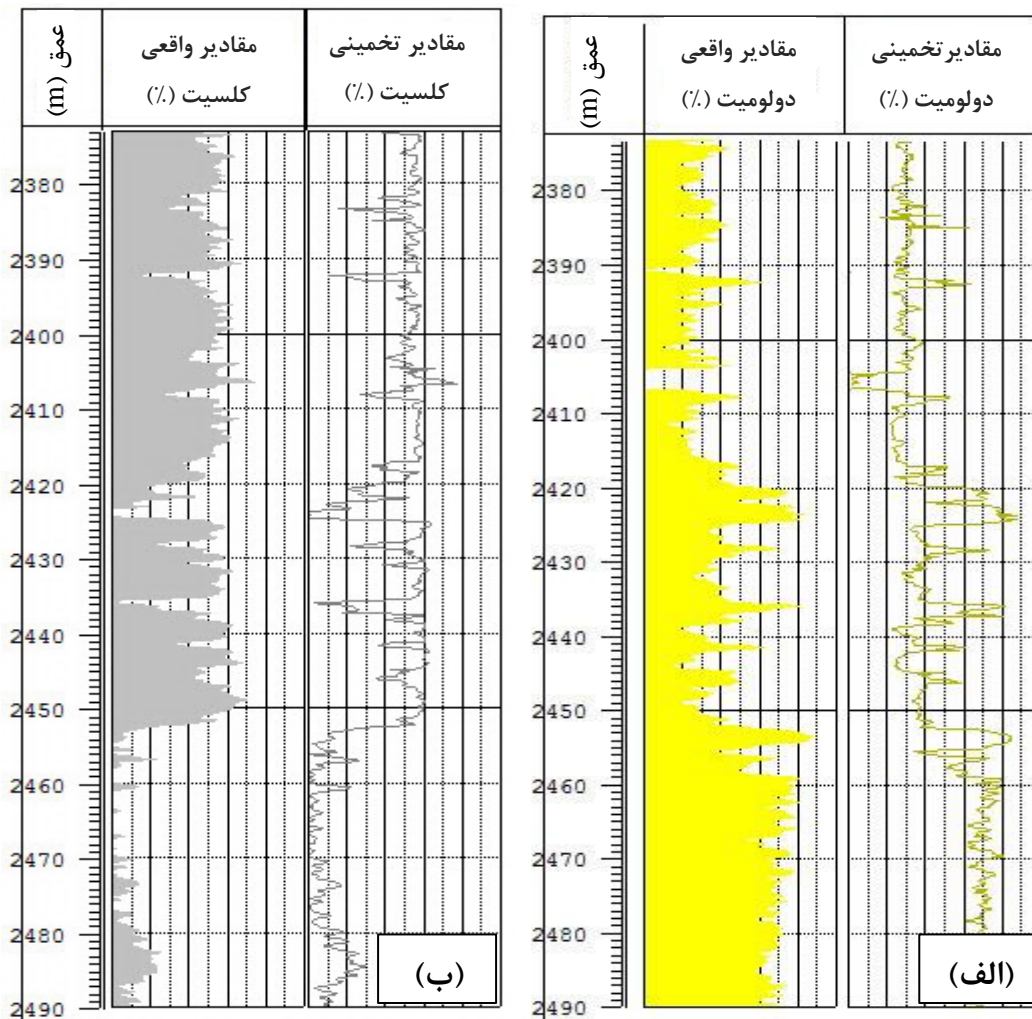
میانگین فاصله کل نمونه موجود در خوشه‌های دیگر با یک خوشه مشخص تعریف می‌شود. در روش‌های کالینسکی - هاراباسز، کرزانوسکی - لای از مجموع مربعات پراش بین خوشه‌ها و درون خوشه‌ای استفاده می‌شود. به عنوان مثال برای محاسبه‌ی شاخص کرزانوسکی - لای، ابتدا تابع تفاوت بر اساس ماتریس مجموع مربعات درون خوشه‌ای تعریف می‌شود و نسبت تابع تفاوت برای خوشه‌ها دو به دو بررسی می‌شود. همچنین معیار دیویس - بولدین تابعی از نسبت مجموع پراش درون خوشه به فاصله بین خوشه‌ها است. در سه معیار نخست هرچه مقدار به دست آمده بزرگتر باشد، نتیجه مطلوب‌تر خواهد بود. اما در شاخص دیویس - بولدین عکس آن صادق است و مقادیر کمتر نشان‌دهنده‌ی خوشه‌بندی بهتر است. پس از محاسبه‌ی مقدار هر یک از شاخص‌ها، نرمال کردن مقادیر و هم‌ارزش کردن آنان، برای هر خوشه‌ی مشخص جمع چهار شاخص محاسبه شده و خوشه‌ای که در آن جمع شاخص‌ها بیشترین مقدار را داشت به عنوان تعداد خوشه‌ی بهینه انتخاب شد. مطابق جدول ۲، بیشترین مقدار مربوط به تعداد خوشه‌ی ۲ است که تاییدی بر انتخاب ۲ خوشه بر اساس اطلاعات مغزه نیز بود.

جدول ۱: استفاده از شاخص‌های اعتبارسنجی جهت تعیین تعداد خوشه‌ی بهینه

شاخص اعتبارسنجی / تعداد خوشه‌ها	۲	۳	۴	۵	۶
سیلهوته	۰/۴۵۵۱۰۱	۰/۱۸۴۷۱۶	۰/۳۷۷۵۶۴	۰/۰۴۸۴۱۲	۰/۱۰۳۷۶۹
سیلهوته - نرمال شده	۱	۰/۳۳۵۱۵۵	۰/۸۰۹۳۴۶	۰	۰/۱۳۶۱۱۶
دیویس - بولدین	۱/۲۲۶۶۳۴	۱/۷۴۵۴۵۴	۱/۲۸۲۱۴۸	۲/۲۳۶۷۶۴	۱/۷۹۹۵۵۶
دیویس - بولدین - نرمال شده	۱	۰/۴۸۶۳۸۳	۰/۹۴۵۰۴۳	۰	۰/۴۳۲۸۲۳
کالینسکی - هاراباسز	۴۵/۸۶۶۶	۱۳۹/۲۱۰۱	۲۶۲/۰۸۹۲	۱۲۵/۸۰۶۹	۱۹۵/۸۳۸
کالینسکی - هاراباسز - نرمال شده	۱	۰/۰۶۰۹۰۷	۰/۶۱۹۲۹۷	۰	۰/۳۱۸۲۳۷
کرزانوسکی - لای	۴۵/۳۶۸۱۵	۵۳/۷۸۱۱۶	۳۷/۴۹۵۷	۴۹/۲۵۳۳۲	۳۶/۵۴۸۷۷
کرزانوسکی - لای - نرمال شده	۰/۵۱۱۷۹۱	۱	۰/۰۵۴۹۵	۰/۷۳۷۲۴۸	۰
جمع چهار شاخص	۳/۵۱۱۷۹۱	۱/۸۸۲۴۴۵	۲/۴۲۸۶۳۶	۰/۷۳۷۲۴۸	۰/۸۸۷۱۷۶

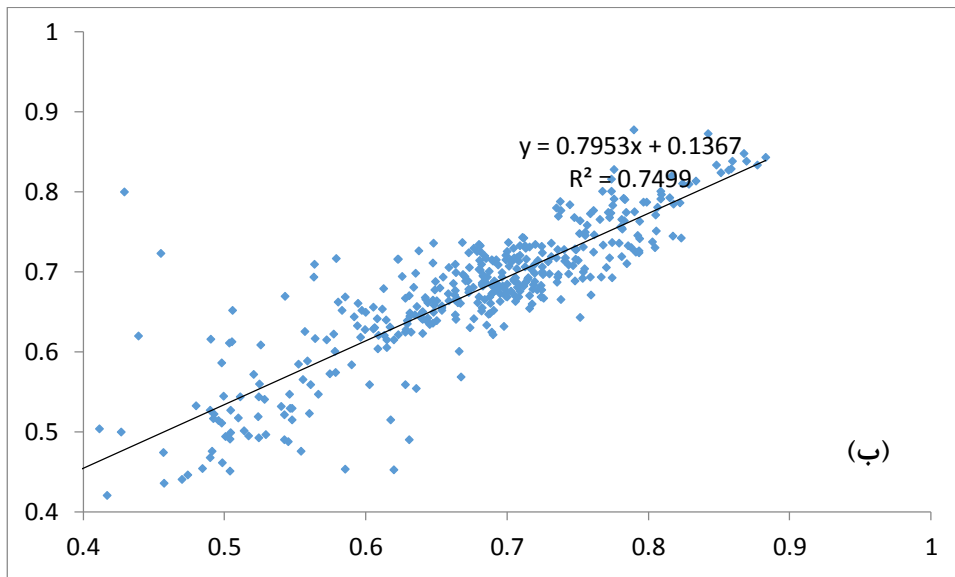
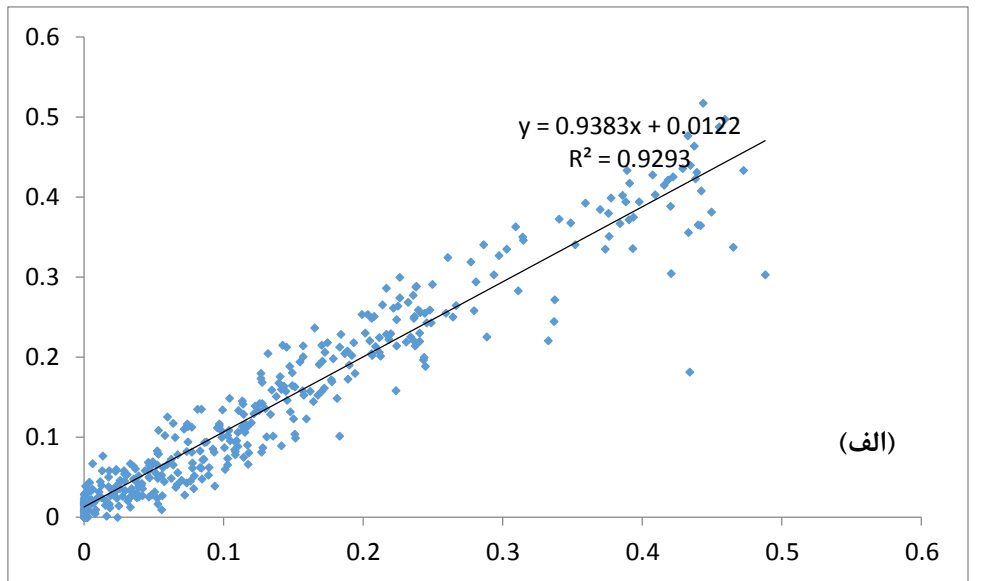
۵. نتایج تخمین بر روی داده‌های واقعی

پس از انجام خوشه‌بندی و تعیین دو جامعه‌ی غالب لیتولوژیکی کلسیت و دولومیت، در این مرحله درصد کانی‌ها در هر یک از جوامع توسط تخمین‌گر MLP تخمین زده شد و با مقادیر به دست آمده از مغزه‌ها مقایسه شدند. نتایجی که در شکل ۷-الف و ۷-ب آورده شده است، به ترتیب مربوط به مقادیر تخمینی دولومیت و کلسیت و مقایسه‌ی آنها با مقادیر واقعی است.



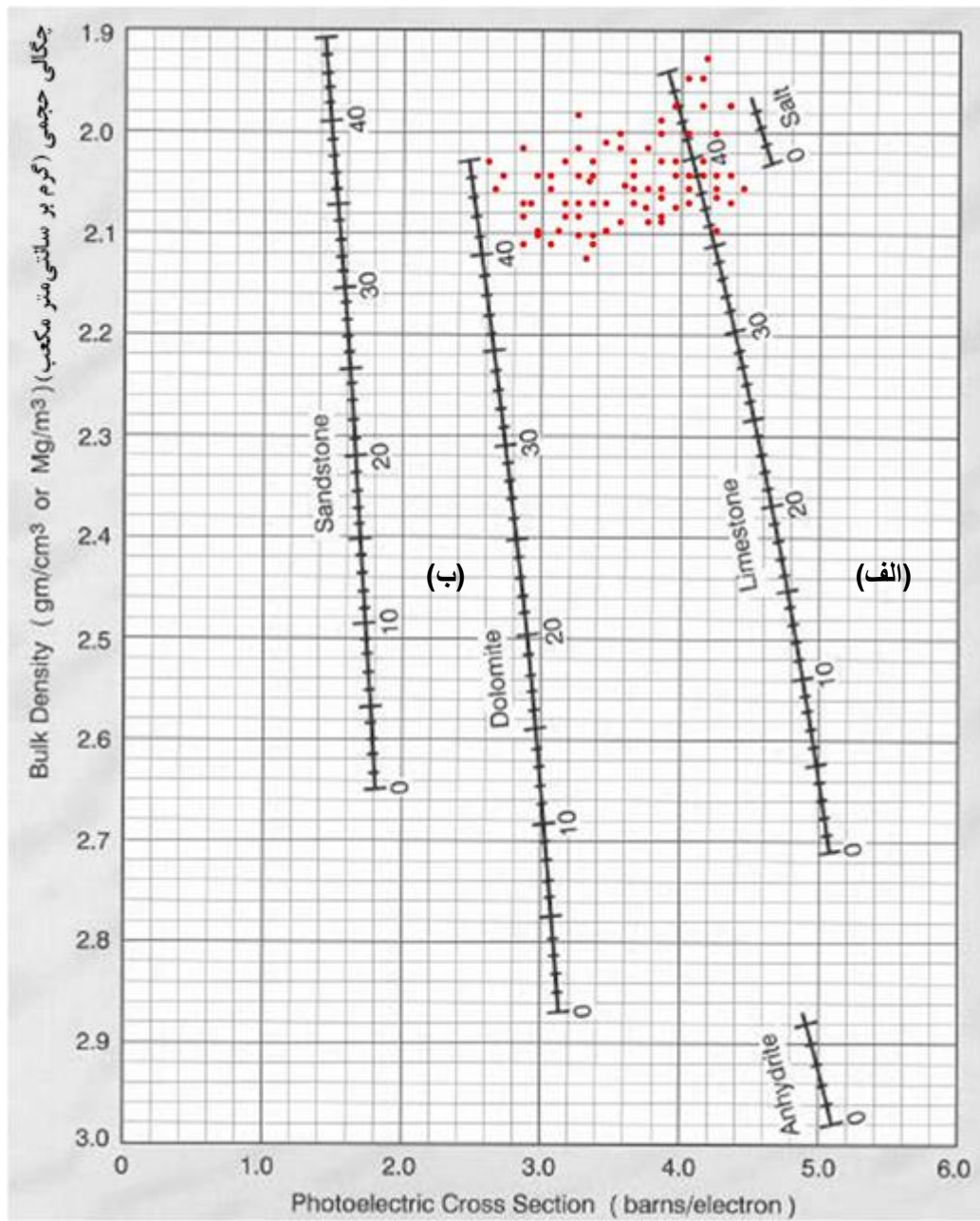
شکل ۷: مقایسه‌ی نتایج تخمین زده شده و واقعی در چاه شماره‌ی ۷، (الف) دولومیت (ب) کلسیت

همچنین در شکل ۸، این همبستگی به صورت دیگری نمایش داده شده است. در این شکل محور افقی مقادیر واقعی کلسیت و محور قائم مقادیر تخمینی هستند. هر چه میزان شیب خط در این نمودار نزدیک به عدد یک باشد، تخمین بهتری صورت گرفته، چرا که هر چه زاویه‌ی خط برازش داده شده بر نقاط نزدیک به ۴۵ درجه باشد به این معنی است که عرض و طول نقاط تقریباً مقادیر یکسانی داشته‌اند و این به معنای تخمین بهینه است. شکل ۸-الف نمایانگر مقادیر تخمینی و واقعی دولومیت در جامعه‌ی کلسیت غالب و شکل ۸-ب نمایانگر مقادیر تخمینی و واقعی کلسیت در جامعه‌ی کلسیت غالب است. همان‌طور که ملاحظه می‌شود در ضریب همبستگی بین مقادیر واقعی و تخمینی در شکل ۸-الف و ۸-ب به ترتیب برابر ۹۲/۹۳٪ و ۷۴/۹۹٪ است.



شکل ۸: (الف) مقادیر تخمینی دولومیت، (ب) مقادیر تخمینی کلسیت

همچنین یکی دیگر از معیارهای سنجش اعتبار در ارتباط با میزان عملکرد تخمین گر، استفاده از نمودارهای استاندارد پتروفیزیکی شلومبرژه است. مطابق شکل ۹ این نمودارهای دو بعدی که هر بعد آن را یکی از ویژگی های پتروفیزیکی تشکیل داده، راهی برای تعیین لیتولوژی ها و مقایسه نتایج به دست آمده از تخمین با مقادیر واقعی هر پارامتر است. پس از تخمین درصدها و تعیین نوع لیتولوژی در طول یک چاه، می توان نقاط را بر روی نمودارهای استاندارد وارد کرد و بررسی نمود که آیا لیتولوژی حاصل از تخمین با لیتولوژی های استاندارد هم خوانی دارد یا خیر.



شکل ۹: نمودار استاندارد چگالی حجمی - اثر فتوالکتریک و قرارگیری نقاط بین دو خط کلسیت و دولومیت

برای مثال در شکل ۹ نقاط تخمین زده شده بر روی نمودار چگالی حجمی - اثر فتوالکتریک قرار داده شده و همان‌طور که دیده می‌شود نقاط در کنار خطوط استاندارد کلسیت، دولومیت و یا در بین این دو قرار گرفتند و این موضوع تطابق کامل با بررسی‌های قبلی را نشان می‌دهد.

نتیجه گیری

استفاده از روش‌های شناسایی الگو راهی مناسب جهت تعیین لیتولوژی و تخمین درصد کانی‌های مخزنی در چاه‌های بدون مغزه است. با بررسی‌های اولیه و آزمایش بر روی داده‌ها، مشاهده شد که عدم اعمال خوشه‌بندی در مرحله‌ی نخست موجب بروز و بالا رفتن خطا در مرحله‌ی تخمین می‌شود، لذا با انتخاب روش خوشه‌بندی مناسب با توجه به ساختار داده‌ها در جهت تعیین لیتولوژی‌های غالب، داده‌ها ابتدا در گروه‌هایی قرار گرفتند که با سایر داده‌های هم‌گروه خود تشابه بیشتری داشتند و خصوصیات پتروفیزیکی آنها مشابه یکدیگر بود. در ادامه از الگوریتم MLP، که یکی از روش‌های شناخته شده‌ی شناسایی الگو است، جهت تخمین درصد کانی‌ها استفاده شد. نتایج به دست آمده با درصد‌های حاصل شده از مغزه‌ها مقایسه شد و مشاهده شد که تخمین‌گر قادر به تولید ضریب همبستگی $۰.۹۲/۹۳$ و $۰.۷۴/۹۹$ بین داده‌های واقعی و تخمینی دولومیت و کلسیت است. با به دست آمدن این نتایج، می‌توان انتظار داشت که این الگوریتم‌ها در چاه‌هایی که فاقد مغزه‌ی حفاری هستند نیز پاسخ‌های قابل توجهی خواهند داد و بدون نیاز به صرف هزینه‌های زیاد برای تهیه مغزه، به کمک این روش‌ها و همچنین اطلاعات کلی که از منطقه موجود است، می‌توان به تخمین پارامترهای مورد نیاز پرداخت.

"هیئت تحریریه مجله از آقایان دکتر عرب امیری و مهندس محمد محمدنیا که داوری مقاله را بر عهده داشته اند کمال تشکر و سپاس را دارد"

منابع

- [1] Delfiner P., O. Peyret and O. Serra, 1987, Automatic Determination of Lithology From Well Logs, SPE Formation Evaluation, Volume 2, Number 3.
- [2] Chang H., D. Kopaska-Merkel and Chen, 2000, Lithofacies Identification using Multiple Adaptive Resonance Theory Neural Networks and Group Decision Expert System, Computers and Geosciences p591 – 601.
- [3] Schlumberger, 1991, Log Interpretation Charts.
- [4] Egbai J. C. and C. O. Aigbogun, 2012, Mathematical modelling of petrophysical parameters for reservoir characterization using well log data, Science Research, 656-670.
- [5] Akinyokun O. C., P. A. Enikanselu and A. B. Adeyemo, 2009, Well Log Interpretation Model for the Determination of Lithology and Fluid Contents, The Pacific Journal of Science and Technology, Volume 10, Number 1.
- [6] Serir L., E. Ramasso and N. Zerhouni, 2012, Evidential Evolving Gustafson-Kessel Algorithm For Online Data Streams Partitioning Using Belief Function Theory, Symbolic and Quantitative Approaches to Reasoning with Uncertainty Lecture Notes in Computer Science Vol. 6717.
- [7] <http://ceit.aut.ac.ir/~shiry/lecture/machine-learning/tutorial/>
- [8] Awodele O. and O. Jegede, 2009, Neural Networks and Its Application in Engineering, Proceedings of Informing Science & IT Education Conference (InSITE).
- [9] Gurney K., 1997, An Introduction to Neural Network, 148 Pages.
- [10] Baan M. and Ch. Jutten, 2000, Neural networks in geophysical applications, Geophysics, Vol. 65, NO. 4, P. 1032-1047.
- [11] Veulenturf L. P. J., 1995, Analysis and Application of Neural Networks.
- [۱۲] منهاج، م. ب.، ۱۳۸۴، مبانی شبکه‌های عصبی، نشر دانشگاه صنعتی امیرکبیر، جلد اول، چاپ سوم، تهران.
- [13] Houkjær K., K. Torp and R. Wind, 2006, Simple and Realistic Data Generation.
- [14] Kovács F., C. Legány and A. Babos, 2006, Cluster Validity Measurement Techniques, AIKED'06 Proceedings of the 5th WSEAS International Conference on Artificial Intelligence, Knowledge Engineering and Data Bases, Pages 388-393.

Making the lithology log by Clustering-Estimation approach in the wells of an oil field in South of Iran

Seddigh, S., Alipoor Shahsavari, M., Moamerian, H., Tokhmchi, B.

Abstract

Reservoir lithology determination is one of the main studies used for well correlation and analyzing productive zones of the reservoir. The best way for lithology determination is using core and cutting information. Nevertheless, in most wells these data is not complete and continual, so in these cases usually use well logging for lithology estimation. The purpose of this paper is representing accurate method for lithology estimation of petrophysical well data with Clustering-Estimation approach. This method has been generalized according to one well from one of the oil fields in South of Iran that contains core data. Then this method is generalized in uncored wells. Clustering is used as a way for grouping well data in homogeneous lithology clusters. Afterward, percentage of mineral is estimated in each of these clusters. The regression coefficients are calculated 92.93% and 74.99% between real and estimated data respectively for calcite and dolomite in one of the wells. The results with high accuracy show the generalization of this method.

Keywords: Asmari reservoir, Clustering, Lithology, MLP estimator, K-means clustering, Gustafson-Kessel clustering, Iran.