

مقایسه روش‌های شبکه عصبی خود سازنده و آنالیز خوشه‌ای برای ارزیابی مقدار کربن آلی در سازندهای محتوی هیدروکربن با استفاده از سیستم‌های هوشمند

پژوهش‌نفت

سال بیست و سوم

شماره ۷۵

صفحه ۱۱۷-۱۳۰ ۱۳۹۲

تاریخ دریافت مقاله: ۹۰/۹/۱۲

تاریخ پذیرش مقاله: ۹۱/۲/۲۶

ابراهیم سفیداری^{۱*}، علی کلخدانی^۲ و محمد شریفی^۱

۱. دانشکده زمین‌شناسی، پردیس علوم، دانشگاه تهران

۲. گروه زمین‌شناسی، دانشکده علوم طبیعی، دانشگاه تبریز

ebrahimspidari@ut.ac.ir

واحدهای مشخص (الکتروفاسیس) در مقایسه با مدل استخراج شده برای کل مجموعه داده‌ها بدون در نظر گرفتن خوشه‌بندی، مقدار کل کربن آلی سازنده را با دقت بالاتری پیش‌بینی می‌نماید. در مجموع سیستم‌های هوشمند نسبت به تکنیک‌های قدیمی مبتنی بر روش $\Delta \log R$ مناسب‌تر می‌باشند. روش ارائه شده همراه با مثال موردی از بزرگ‌ترین مخزن گازی غیر همراه جهان، میدان گازی پارس جنوبی در حوضه خلیج فارس ارائه گردیده است.

واژه‌های کلیدی: خوشه‌بندی SOM، آنالیز خوشه‌ای، شبکه عصبی خود سازنده، الکتروفاسیس، ارزیابی اعتبار خوشه‌بندی

مقدمه

مقدار کل کربن آلی یکی از پارامترهای مهم در ارزیابی ژئوشیمیایی سنگ منشاءهای حاوی هیدروکربن می‌باشد. افزایش مقدار TOC^۱ احتمالاً نشان‌دهنده حضور سنگ

چکیده

محتوای کل کربن آلی یکی از پارامترهای مهم جهت ارزیابی ژئوشیمیایی لایه‌های تولید کننده نفت و گاز است. در این مطالعه، طی دو مرحله، محتوای کربن آلی در سازندهای هیدروکربن‌دار با استفاده از داده‌های لاگ ارزیابی شده است. در مرحله اول، داده‌های لاگ به مجموعه‌ای از الکتروفاسیس‌ها تقسیم‌بندی شده‌اند. روش‌های استفاده شده برای شناسایی و خصوصیت‌بندی الکتروفاسیس‌ها شامل: شبکه‌های عصبی خود سازنده و روش آنالیز خوشه‌ای می‌باشد. نتایج حاصل از هر دو روش با یکدیگر مقایسه شده و براساس آزمون‌های ارزیابی خوشه‌ای، بهترین روش برای خوشه‌بندی داده‌های پتروفیزیکی در الکتروفاسیس‌های معین مورد استفاده قرار گرفت. مقدار کل کربن آلی با استفاده از داده‌های لاگ به وسیله روش‌های خاص شبکه عصبی برای هر الکتروفاسیس برآورد شد. در مرحله دوم، مقدار کل کربن آلی با استفاده از همان روش خاص شبکه عصبی و بدون در نظر گرفتن الکتروفاسیس‌ها تعیین گردید. نتیجه دو روش با یکدیگر و همچنین با روش $\Delta \log R$ مقایسه شد. نتایج نشان داد که خوشه‌بندی یک سازنده به

1. Total Organic Carbon

در معادلات بالا، عدد یک مشاهده یا لاگ اندازه‌گیری شده در یک عمق خاص، x نماینده یک لاگ خاص و p شماره تعداد لاگ‌ها می‌باشد. عمومی‌ترین روش اندازه‌گیری فاصله بین دو نمونه بردار روش فاصله اقلیدوسی و فاصله اقلیدوسی استاندارد می‌باشد که روابط مورد استفاده در این دو روش به صورت زیر است:

$$d_{rs}^2 = (x_r - x_s)(x_r - x_s) \quad (3)$$

فاصله استاندارد:

$$dS_{rs}^2 = (x_r - x_s) D^{-1} (x_r - x_s) \quad (4)$$

در روابط بالا d فاصله اقلیدوسی، ds مشاهدات، r و s شماره مشاهدات، D ماتریس قطری با عناصر مورب و $v2$ واریانس x متغیر در m مشاهده می‌باشد. آنالیز خوشه‌ای پایکانی شامل سه مرحله به شرح زیر است:

- ۱- پیدا کردن شباهت یا اختلاف بین هر جفت از مشاهدات موجود در مجموعه داده‌ها
- ۲- گروه‌بندی کردن مشاهدات به صورت درخت خوشه‌ای پایکانی
- ۳- مشخص کردن محل برش بر روی درخت پایکانی

در این مطالعه داده‌های ورودی اصلی آنالیز خوشه‌ای شامل لاگ‌های کنترل‌کننده لیتولوژی و تخلخل شامل NPHI، DT RHOB، GR، CGR، SGR، URAN، THOR و POTA می‌باشد. افزایش میزان هتروژنی مخزن، افزایش تعداد کلاسترها را در پی خواهد داشت. نتایج نشان می‌دهد که تشخیص ۷ خوشه بیشترین تناسب را با مطالعات زمین‌شناسی اولیه در منطقه دارد [۶]. شکل ۱ نتایج خروجی آنالیز خوشه‌ای را نشان می‌دهد.

شبکه عصبی خود سازنده^۳ (SOM)

SOM یکی از مهم‌ترین مدل‌های شبکه عصبی است که اخیراً در زمینه‌های مختلفی از قبیل آنالیز تصاویر [۷ و ۸] جهت حرکت بادها [۹ و ۱۰] و تفسیر امواج لرزه‌ای [۱۱] و [۱۲] به کار رفته است.

منشاء و کاهش آن نشان دهنده کاهش گسترش سنگ منشاء می‌باشد. پیرولیز راک ایول^۱ یکی از روش‌های استاندارد است که به صورت گسترده برای ارزیابی پتانسیل زایشی هیدروکربن مورد استفاده قرار می‌گیرد [۱]. با این وجود آنالیزهای ژئوشیمیایی اغلب بر هزینه و زمان بر است. از این رو توسعه روشی سریع با هزینه کم برای پیش‌بینی داده‌های ژئوشیمیایی با صحت قابل قبول بسیار مهم می‌باشد. از طرف دیگر داده‌های پتروفیزیکی برای همه چاه‌های حفاری شده در یک میدان هیدروکربنی در دسترس است. در این زمینه محققان زیادی درخصوص رابطه بین داده‌های ژئوشیمیایی و پتروفیزیکی تحقیق کرده‌اند [۲-۵]. به دلیل پیچیدگی رابطه بین داده‌های پتروفیزیکی و ژئوشیمیایی، مطالعات اخیر بر روی سیستم‌های هوشمند متمرکز شده و نتایج قابل قبولی نیز به دست آمده است. با گسترش محدوده داده‌ها، میزان خطا در سیستم‌های هوشمند افزایش خواهد یافت. از این رو مطالعه حاضر سعی می‌نماید به وسیله الکتروفاسیس‌ها^۲ و استخراج مدل‌های هوشمند برای هر الکتروفاسیس منحصر به فرد، میزان خطا را کاهش دهد.

روش کار

تعیین الکتروفاسیس‌ها بر مبنای شناسایی گروه داده‌هایی با خصوصیات یکسان در داده‌های حاصل از لاگ‌ها انجام می‌گیرد. در ادامه، سه روش برای شناسایی الکتروفاسیس‌ها مورد بحث قرار گرفته است.

آنالیز خوشه‌ای

هدف از آنالیز خوشه‌ای تقسیم‌بندی مجموعه داده‌ها به گروه‌های مشخص بر اساس اندازه‌گیری شباهت یا تفاوت بین گروه‌ها می‌باشد که داده‌های موجود در هر گروه بیشترین شباهت را با یکدیگر و بیشترین تفاوت را با داده‌های موجود در گروه‌های دیگر دارند. الگوریتم‌های مختلفی برای محاسبه فاصله بین دو داده‌برداری وجود دارد. اگر دو مجموعه داده مشخص، شامل اندازه‌گیری‌های چندگانه‌ای در یک مشاهده یکسان باشند، می‌توانند به وسیله بردارهای چند بعدی زیر تعریف شوند:

$$X1 = (x11 \ x12 \ \dots \ x1p) \quad (1)$$

$$X2 = (x21 \ x22 \ \dots \ x2p) \quad (2)$$

1. Rock-Eval Pyrolysis
2. Electrofacies
3. Self Organizing Map

کلاس‌های بهینه مورد استفاده قرار می‌گیرد که این میزان بر اساس دوری و نزدیکی مشاهدات و خوشه‌ها به یکدیگر محاسبه می‌شود. مقدار $S(i)$ با استفاده از فرمول زیر قابل محاسبه است:

$$S(i) = \frac{(b(i) - a(i))}{\max \{a(i), b(i)\}} \quad (5)$$

$a(i)$ میانگین فاصله بین مشاهده i با سایر مشاهدات در یک خوشه مشابه و $b(i)$ میانگین فاصله مشاهده i به تمام مشاهدات در خوشه‌های دیگر می‌باشد. بر اساس فرمول بالا مقدار $S(i)$ بین -1 و $+1$ قرار دارد. اگر $S(i)$ به $+1$ نزدیک‌تر باشد، به این معنی است که خوشه‌بندی نمونه خوب صورت گرفته است و خوشه پیشنهاد شده برای نمونه مورد نظر مناسب می‌باشد. ولی اگر $S(i)$ به -1 نزدیک‌تر باشد به این معنی است که خوشه‌بندی نمونه، به خوبی انجام نشده و خوشه پیشنهاد شده برای داده مورد نظر نامناسب می‌باشد. بیشترین مقدار $S(i)$ کلی، نشان‌دهنده بهترین روش خوشه‌بندی می‌باشد. همچنین بهترین شماره تعداد خوشه زمانی است که بیشترین مقدار میانگین $S(i)$ کلی را برای کل مجموعه داده‌ها داشته باشیم. مقدار میانگین $S(i)$ برای هر دو روش خوشه‌بندی در شکل ۳ نشان داده شده است. مطابق این شکل، روش شبکه عصبی نتیجه بهتری برای مجموعه داده‌های مورد استفاده در این تحقیق به دنبال دارد.

شبکه عصبی و الگوریتم آموزشی

مدل شبکه عصبی از دو مرحله تشکیل شده است:

- ۱- انتخاب ورودی مناسب از میان مجموعه داده‌های لاگ در دسترس.
- ۲- طراحی شبکه عصبی پس انتشار خطا با الگوریتم آموزشی مناسب. واضح است که بخش‌هایی از روش شرح داده شده در بالا در تحقیق‌های گذشته هم وجود داشته و جدید نمی‌باشد. به عنوان مثال الگوریتم آموزشی شبکه عصبی روش مرسوم است. اختلاف بین مطالعه حاضر با کارهای قبلی در استفاده از برخی الگوهای اضافی مبتنی بر مشخص کردن الکتروفاسیس‌ها برای برآورد مناسب‌تر و قابل قبول تر TOC بر اساس لاگ‌های ورودی می‌باشد.

الگوریتم SOM یک نوع مدل شبکه عصبی است که در پیاده‌سازی و طرح ریزی مشخصه‌های غیر خطی از فضای چند بعدی به فضای یک بعدی مورد استفاده قرار می‌گیرد [۱۳]. الگوریتم آموزشی شبکه‌های عصبی خود سازنده به صورت زیر خواهد بود [۱۴-۱۶].

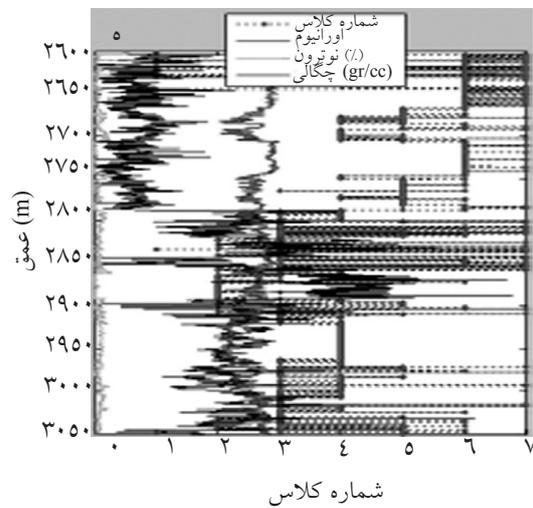
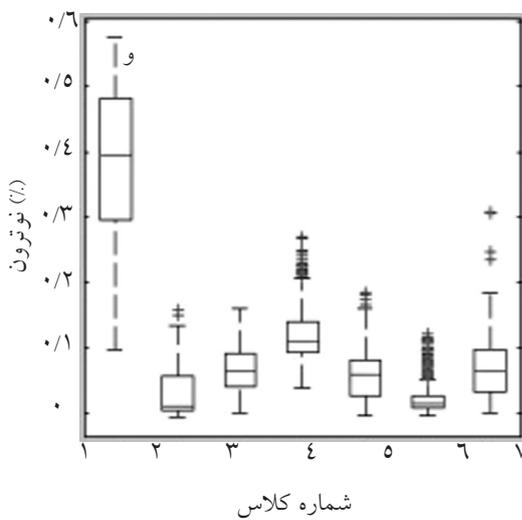
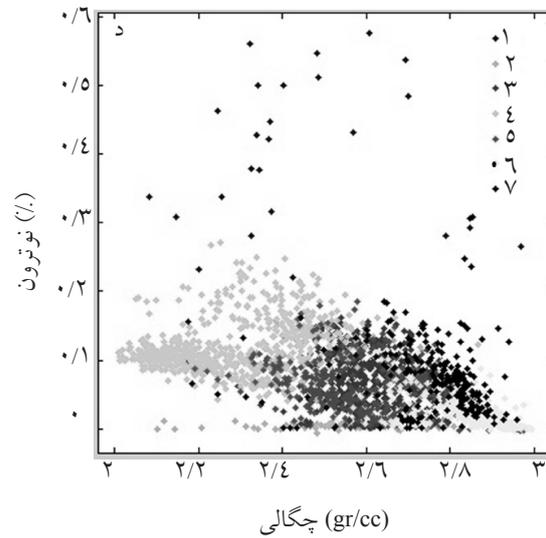
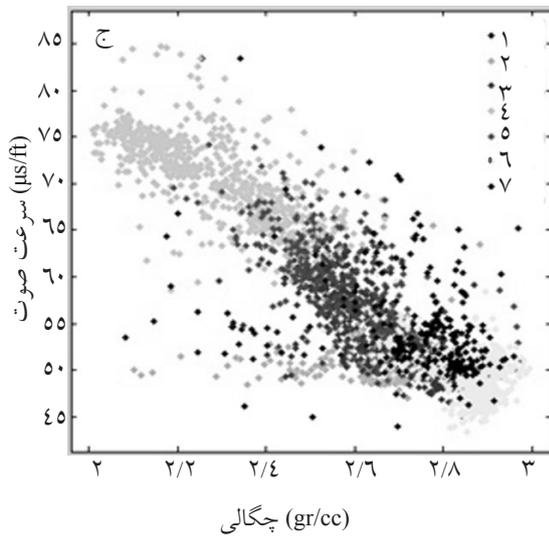
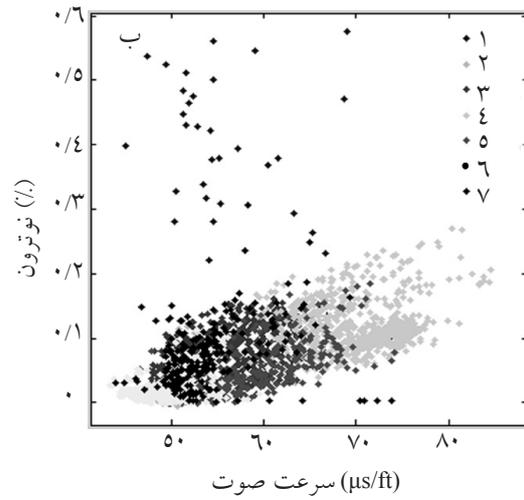
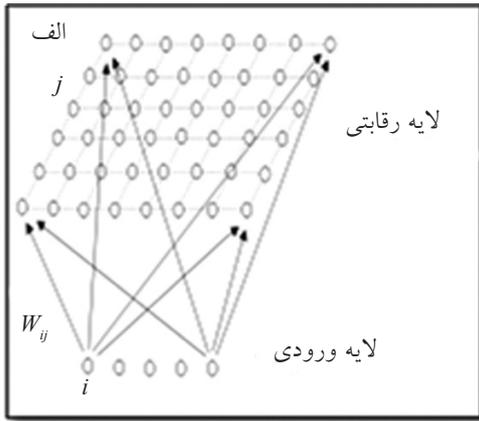
۱. محاسبه فاصله بین الگو و تمام سلول‌های عصبی ($d_{ij} = \|x_k - w_{ij}\|$)
۲. انتخاب نزدیک‌ترین نرون به عنوان نرون برنده ($d_{ij} = \min(d_{mn})$)
۳. به‌روزرسانی هر نرون با توجه به قاعده $w_{ij} = w_{ij} + \alpha h(w_{winner}, w_{ij}) \|x_k - w_{ij}\|$
۴. این روند تا زمانی که یک معیار توقف خاص به دست آید، تکرار می‌شود. معمولاً معیار توقف، تعداد ثابتی از تکرار می‌باشد. برای اثبات همگرایی و ثبات نقشه، نرخ یادگیری و شعاع همسایگی در هر تکرار کاهش می‌یابد. بنابراین، همگرایی به سمت صفر میل خواهد کرد. فاصله اندازه‌گیری بین بردارها فاصله اقلیدوسی می‌باشد، ولی از سایر اندازه‌گیری‌های فواصل مانند فاصله Mahalanobis نیز می‌توان استفاده نمود. دیاگرام شماتیک ساختار SOM و نتایج خوشه‌بندی آن بر روی داده‌های لاگ استفاده شده، در شکل ۲ به نمایش درآمده است.

ارزیابی خروجی خوشه‌بندی

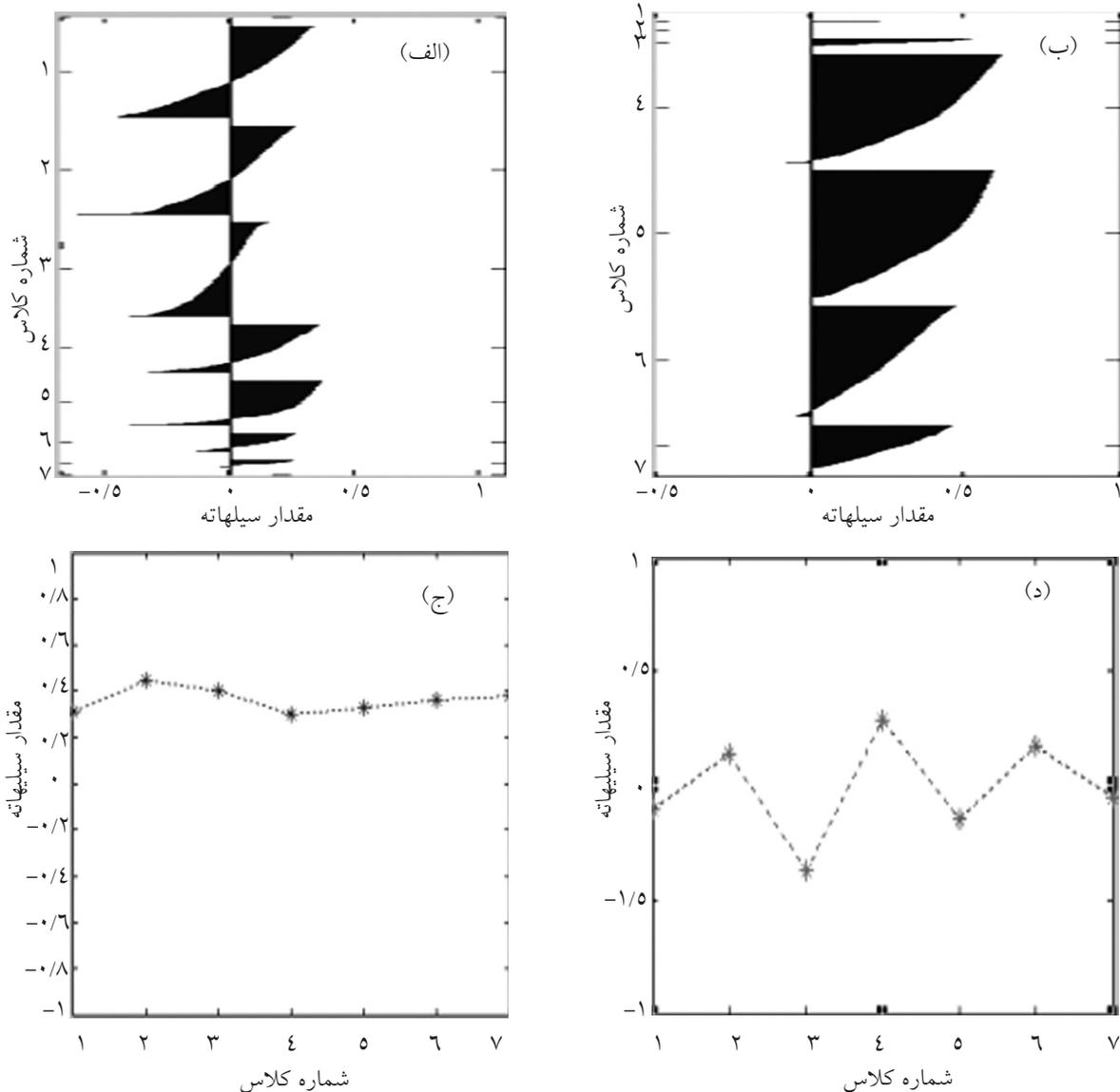
ارزیابی خوشه‌ای اندازه‌گیری، میزان برتری یک خوشه‌بندی نسبت به خوشه‌بندی‌های دیگر به وسیله الگوریتم‌های متفاوت خوشه‌بندی یا الگوریتم‌های مشابه ولی با مقدار پارامترهای متفاوت می‌باشد [۱۷]. اندیس C (C-index) و روش اعتبارسنجی Silhouette^۱ در این مطالعه برای مقایسه روش‌های خوشه‌بندی مورد استفاده قرار گرفته است.

روش اعتبارسنجی Silhouette

تکنیک اعتبارسنجی Silhouette مقدار Silhouette را برای هر نمونه، هر کلاس و مجموعه داده‌ها محاسبه می‌کند. میانگین مقدار Silhouette برای ارزیابی اعتبار خوشه‌بندی و همچنین برای تصمیم‌گیری در مورد انتخاب تعداد



شکل ۲- نتایج خروجی حاصل از روش SOM. نمایی از یک شبکه عصبی خود سازمانده (الف)، نمودار تفکیک تخلخل صوتی در مقابل نوترون، چگالی در مقابل سرعت صوت (ج)، چگالی در مقابل نوترون (د)، نمودار جعبه‌ای تفکیک تخلخل نوترون در کلاس‌های مختلف و لاگ‌های اورانیوم، نوترون چگالی و کلاس‌های الکتریکی مختلف در مقابل عمق.



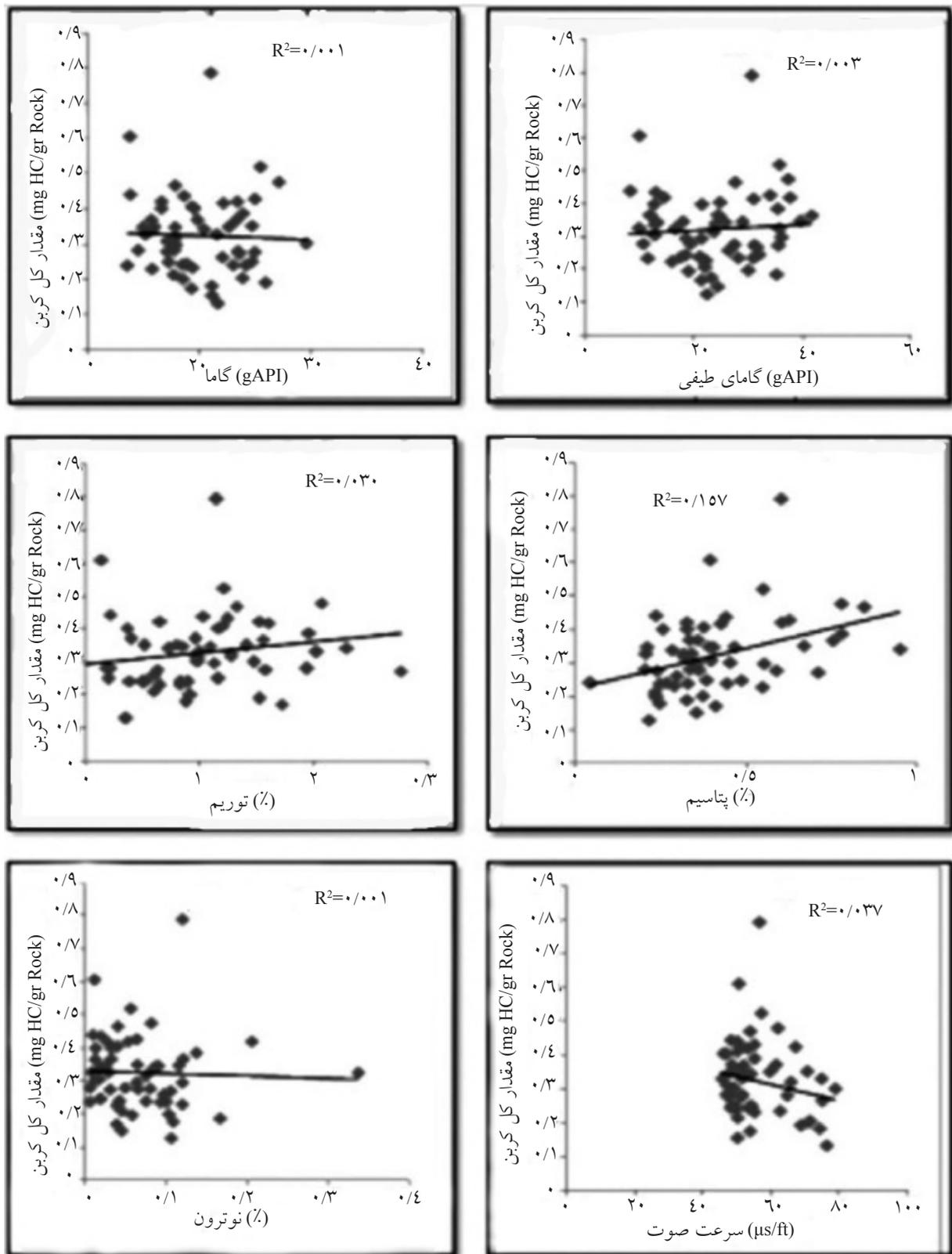
شکل ۳- مقدار سیلهاته برای روش شبکه عصبی (الف و ج) و آنالیز پایکانی (ب و د). همان‌طور که دیده می‌شود برای روش شبکه عصبی این مقدار به ۱ نزدیکتر می‌باشد.

بین داده‌های ورودی و TOC قبل و پس از خوشه‌بندی در شکل ۴ و ۵ نشان داده شده است. همان‌طور که دیده می‌شود، این همبستگی بعد از خوشه‌بندی بهبود یافته و برای الکتروفاسیس شماره ۱ به تصویر کشیده شده است.

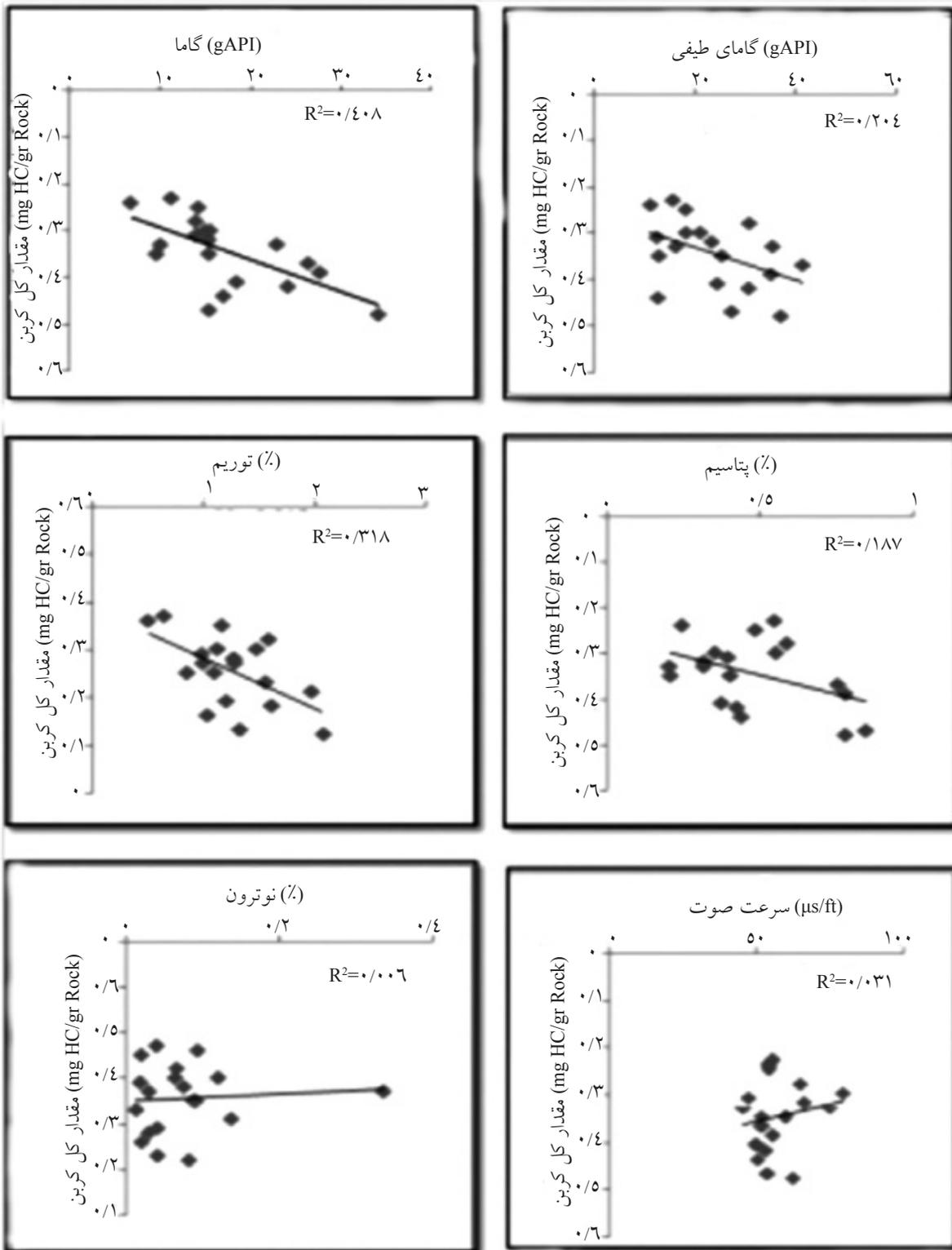
عموماً اندیس هیدروژن (HI) در مواد آلی به دلیل بالا بودن محتوای هیدروژن بالا می‌باشد. بنابراین تخلخل نوترون در بازه‌های غنی از مواد آلی افزایش می‌یابد. زمان انتقال صوت، تابعی از لیتولوژی سازند، تخلخل و مدل‌های توزیع سیال در سازند است (نفت، آب، گاز و ...). مقدار TOC با افزایش مقادیر (DT)، GR، CGR، SGR، پتاسیم و توریم افزایش می‌یابد.

انتخاب ورودی‌های مناسب

این مرحله نقش مهمی در مدل استخراج شده ایفا می‌نماید. مقایسه نشان می‌دهد که لاگ‌های تخلخل نوترون (NPHI)، زمان انتقال صوت (DT)، GR، CGR، SGR، THOR و K بیشترین همبستگی را با TOC دارند. نتایج این مطالعه نشان می‌دهد که بعد از خوشه‌بندی داده‌های پتروفیزیکی، همبستگی بین ورودی‌ها و خروجی‌ها در هر الکتروفاسیس به میزان قابل ملاحظه‌ای افزایش می‌یابد. این نکته باعث اختلاف این مقاله با دیگر کارهای انجام شده و کاهش خطای پیش‌بینی به وسیله خوشه‌بندی بر مبنای یک مدل شبکه عصبی می‌باشد. مقایسه همبستگی



شکل ۴- همبستگی بین داده‌های ورودی و خروجی قبل از خوشه بندی. همبستگی بین مقدار کل کربن آلی با لاگ های ورودی، مقدار کربن آلی - گاما (الف)، مقدار کربن آلی - طیف گاما (ب)، مقدار کربن آلی - توریم (ج)، مقدار کربن آلی - پتاسیم (د)، مقدار کربن آلی - نوترون (ه) و مقدار کل کربن آلی - سرعت صوت (و).



شکل ۵- همبستگی بین داده‌های ورودی و خروجی بعد از خوشه بندی. همبستگی بین مقدار کل کربن آلی با لاگ های ورودی، مقدار کربن آلی - گاما (الف)، مقدار کربن آلی - طیف گاما (ب)، مقدار کربن آلی - توریم (ج)، مقدار کربن آلی - پتاسیم (د)، مقدار کربن آلی - نوترون (ه) و مقدار کل کربن آلی - سرعت صوت (و).

به منظور انتخاب ورودی‌های مناسب برای طراحی شبکه عصبی با یک شبکه سه لایه با پارامتر خروجی TOC از نرم‌افزار MATLAB استفاده شد. در لایه ورودی چندین گروه از مجموعه لاگ‌ها (۱۴۰ داده برای آموزش و ۳۰ داده برای اعتبار سنجی) مورد استفاده قرار گرفت. در هر بار اجرا، صحت مدل با داده‌های تست (۳۰ داده) مورد بررسی قرار گرفت.

طراحی شبکه با الگوریتم آموزشی

شبکه پس انتشار خطا یک روش آموزشی نظارتی است که ورودی‌ها را به شبکه انتقال می‌دهد و خطای بین خروجی طراحی شده را با خروجی به دست آمده از مجموعه داده‌های آموزشی مقایسه می‌کند. میزان خطا در شبکه به عقب منتقل شده و بر اساس آن وزن‌ها در طول شماری از تکرار که epochs نامیده می‌شود تعدیل می‌گردد. آموزش زمانی متوقف می‌شود که مقدار خروجی محاسبه شده بیشترین نزدیکی را به خروجی طراحی شده داشته باشد [۱۸]. در ارتباط با روش استفاده شده برای به روزکردن وزن‌ها و مقدار بایاس‌ها چندین الگوریتم آموزشی توسعه داده شده است. در این مطالعه الگوریتم آموزشی Levenberg-Marquardt به عنوان بهترین تابع برای آموزش شبکه انتخاب شد. Levenberg-Marquardt یک تابع آموزشی است که وزن‌ها و بایاس‌ها را بر اساس نرمال‌سازی Levenberg-Marquardt تعیین می‌نماید [۱۹].

نتایج و بحث

بررسی موردی

داده‌های مورد استفاده برای این مطالعه از چاه‌های میدان گازی پارس جنوبی فراهم شده است. قبل از استخراج مدل، جمع‌آوری داده‌های ژئوشیمیایی و لاگ‌های پتروفیزیکی از بازه‌های مطالعه شده مورد نیاز بود. مدل شبکه عصبی به دو صورت مختلف قبل و بعد از خوشه‌بندی اجرا گردید و نتایج با یکدیگر مقایسه شد. مراحل کار در ادامه توضیح داده شده است.

پیرولیز راک - ایول

در این روش مقدار اندکی از سنگ (۷۰ mg) یا زغال

را در دستگاه با یک اتمسفر خنثی (هلیوم یا نیتروژن) حرارت می‌دهند و مقدار هیدروکربن موجود در نمونه (پیک S_1) و مقداری که پس از بلوغ آزاد شده (پیک S_2) را اندازه‌گیری نمایند. مقدار T_{max} به عنوان یک پارامتر استاندارد نشان‌گر دمایی است که در آن پیک S_2 به ماکزیمم مقدار خود می‌رسد و به عنوان یک پارامتر بلوغ برای نمونه‌های آلی مورد استفاده قرار می‌گیرد. این پارامتر کیفیت مواد آلی نمونه سنگ را جهت اهداف استخراجی نشان می‌دهد.

آماده‌سازی و پردازش داده‌ها

در این مطالعه ۲۰۰ نمونه همراه با داده‌های لاگ و TOC (به دست آمده از پیرولیز راک ایول) مورد استفاده قرار گرفت. در ابتدا، ۱۴۰ داده برای آموزش، ۳۰ نمونه برای اعتبار‌سازی و ۳۰ نمونه نیز برای تست شبکه مورد استفاده قرار گرفت. داده‌های لاگ مورد بررسی در بازه‌های با دقت کم حذف شدند.

پیش‌بینی مقدار TOC توسط شبکه عصبی

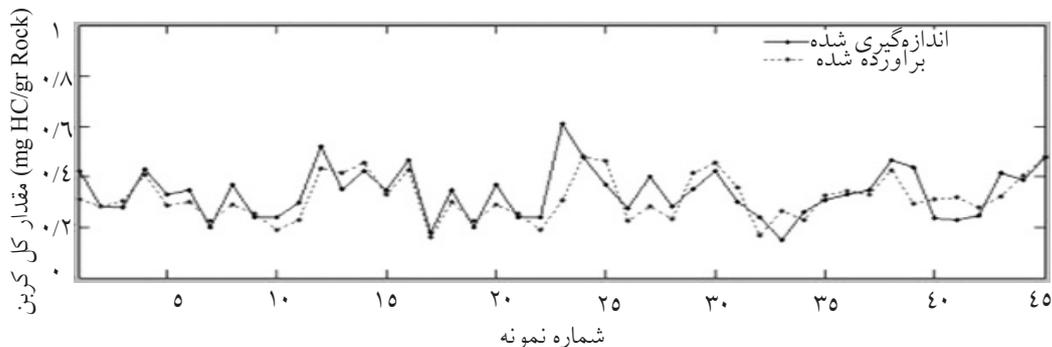
این بخش شامل دو مرحله است: ۱- تخمین مقدار TOC توسط یک شبکه برای کل ایتروال بدون خوشه‌بندی و ۲- تخمین TOC توسط شبکه‌های مجزا برای هر الکتروفاسیس. در ابتدا یک شبکه سه لایه پس انتشار خطا در محیط MATLAB طراحی گردید. برای طراحی شبکه‌ای با بهترین الگوریتم‌های آموزش، مجموعه‌ای از پارامترهای بهینه شامل شماره لایه‌های پنهان، شماره نرون در لایه‌های پنهان، تعداد تکرارهای آموزش و توابع انتقال مورد نیاز است. این پارامترها به روش آزمون و خطا تعیین شدند. برای به دست آوردن مدل بهینه، شبکه‌ای با تعداد مختلفی از نرون‌ها در لایه پنهان به همراه ترکیبی از لاگ‌های ورودی مختلف، مورد آزمایش قرار گرفتند.

مدل‌های شبکه عصبی مجزا برای هر الکتروفاسیس طراحی گردید. پارامترهای مدل بهینه شبکه عصبی که برای هر الکتروفاسیس تعیین شده، در جدول ۲ فهرست شده است. نمایش گرافیکی و پلات‌های تصویری مقایسه ای TOC اندازه‌گیری شده و محاسبه شده قبل و پس از خوشه‌بندی در شکل ۸ نمایش داده شده است. همان‌گونه که دیده می‌شود، نتایج برای الکتروفاسیس اول پس از اعمال خوشه‌بندی، بهبود یافته است.

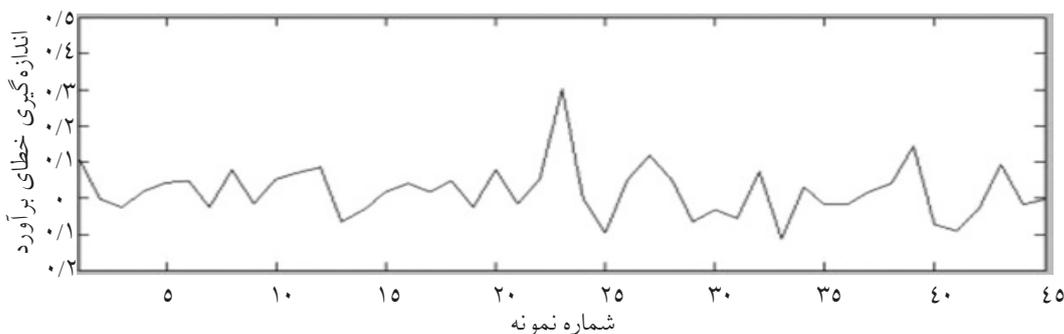
سرانجام مدلی با بالاترین سطح کارایی به عنوان مدل بهینه انتخاب شد. جدول ۱ خصوصیات شبکه‌های مختلف و مدل بهینه را نشان می‌دهد. ملاک کارایی شبکه براساس تطابق ضریب تاثیر بین داده خروجی و داده مورد انتظار و حداقل مقدار خطا انتخاب گردید. خطای اندازه‌گیری شده بین مقدار واقعی TOC و مقدار پیش بینی شده در مقابل تکرارهای تمرین‌دهی در شکل ۶ و ۷ مشاهده می‌شود. به منظور پیش‌بینی TOC براساس مدل الکتروفاسیس،

جدول ۱- خصوصیات شبکه مدل بهینه و میزان خطا برای شبکه استخراج شده

ورودی	تعداد تکرار	تابع آموزشی	میانگین مربعات خطا
GR	۸	LM	۰/۰۸۱
GR, CGR	۷	LM	۰/۰۶۴
GR, CGR, SGR	۹	LM	۰/۰۹۷
GR, CGR, SGR, THOR	۱۱	LM	۰/۰۵۶
GR, CGR, SGR, THOR, POTA	۸	LM	۰/۰۳۲
GR, CGR, SGR, THOR, POTA, URAN	۶	LM	۰/۰۱۱
GR, CGR, SGR, THOR, POTA, URAN, DT	۴	LM	۰/۰۰۹۴
GR, CGR, SGR, THOR, POTA, DT, NPHI	۶	LM	۰/۰۰۷۳
GR, CGR, SGR, THOR, POTA, DT, NPHI	۷	OSS	۰/۰۰۹۴
GR, CGR, SGR, THOR, POTA, DT, NPHI	۸	BR	۰/۰۱۹



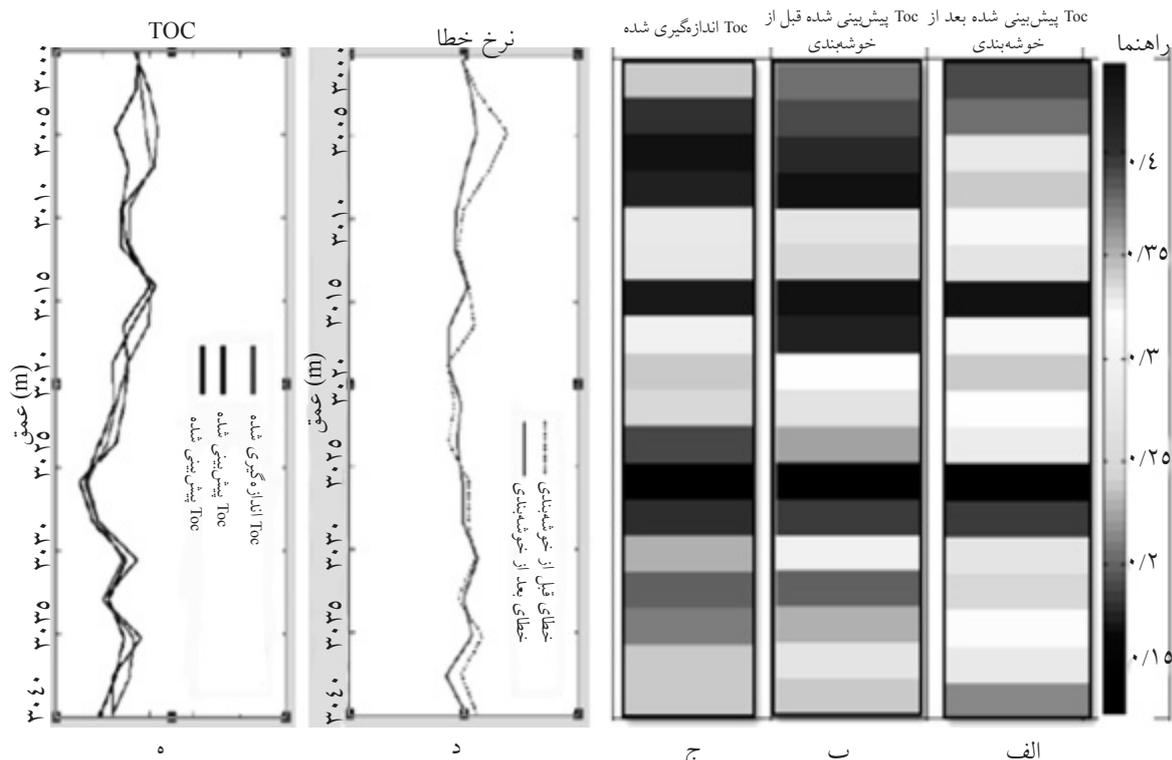
شکل ۶- مقادیر به دست آمده برای خروجی شبکه و داده مورد انتظار در مراحل آموزش، اعتبارسنجی و آزمایش قبل از خوشه‌بندی مقابل شماره نمونه‌ها



شکل ۷- خطای اندازه‌گیری شده بین مقدار واقعی TOC و مقدار پیش‌بینی شده در مقابل تکرارهای مرحله آموزش قبل از خوشه‌بندی مقابل شماره نمونه‌ها

جدول ۲- پارامترهای مدل بهینه شبکه عصبی برای هر الکتروفاسیس

میانگین مربعات خطا	تابع آموزشی	تعداد تکرار	شماره کلاس
۰/۰۰۷۱	LM	۶	۱
۰/۰۰۵۲	LM	۷	۲
۰/۰۰۵۸	LM	۹	۳
۰/۰۰۴۱	LM	۸	۴
۰/۰۰۲۵	LM	۵	۵
۰/۰۰۴۳	LM	۷	۶
۰/۰۰۴۸			میانگین مربعات خطا برای همه کلاس‌ها



شکل ۸- نمایش گرافیکی و پلات‌های تصویری مقایسه‌ای TOC اندازه‌گیری شده و برآورد شده قبل و پس از خوشه‌بندی در الکتروفاسیس ۱. عکس‌های تصویری مقدار کل کربن آلی برآورد شده: بعد از خوشه‌بندی (الف)، قبل از خوشه‌بندی (ب)، مقدار اندازه‌گیری شده (ج)، نرخ خطا برای هر دو حالت (د) و پلات در مقابل عمق برای همه حالت‌های برآورد شده و اندازه‌گیری شده (ه)

$$\Delta \log R = \log_{10} (R/R_{\text{base}}) - k(P - P_{\text{base}}) \quad (6)$$

در این رابطه، R میزان مقاومت و P میزان تخلخل یک لاگ سونیک، دانسیته و یا نوترونی است. R_{base} و P_{base} baseline یا بازه‌های غیرمخزنی هستند. با استفاده از معادله زیر [۲] میزان TOC محاسبه می‌شود:

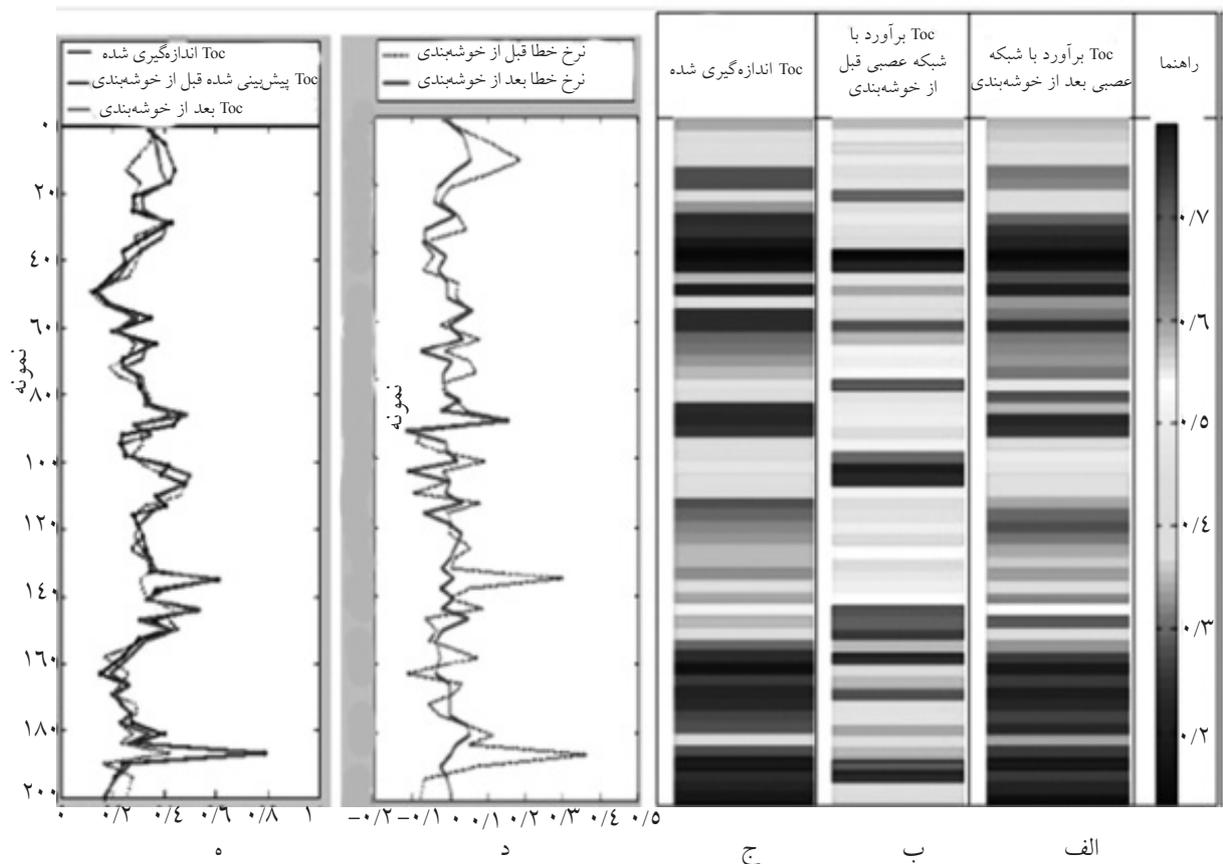
$$\text{TOC} = \Delta \log R * 10 (2.297 - 0.1688 * \text{LOM}) \quad (7)$$

در این رابطه نشان‌گر سطح بلوغ ماده آلی است. LOM‌های بالا (مانند ۱۱) تطابق خوبی را با نتایج آنالیزهای ژئوشیمیایی نشان می‌دهد. مقایسه بین TOC اندازه‌گیری شده حاصل از پیش‌بینی شبکه عصبی و روش $\Delta \log R$ در شکل ۱۰ نمایش داده شده است. به طور کلی، می‌توان گفت که روش شبکه عصبی بسیار دقیق‌تر از روش $\Delta \log R$ است.

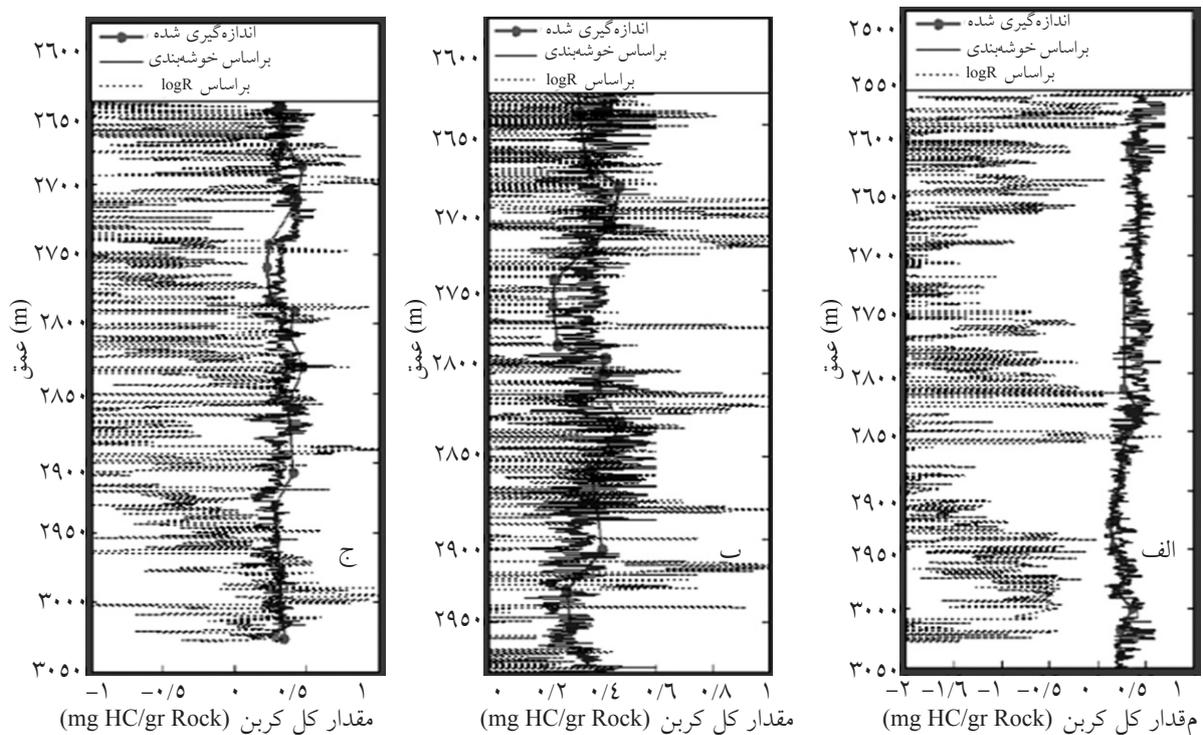
شکل ۸ نمایش گرافیکی و پلات‌های تصویری مقایسه‌ای TOC اندازه‌گیری شده و برآورد شده قبل و پس از خوشه‌بندی را برای کل داده‌ها نشان می‌دهد. شکل ۹ نمایش گرافیکی و پلات‌های تصویری مقایسه‌ای TOC اندازه‌گیری شده و برآورد شده قبل و پس از خوشه‌بندی را برای کل داده‌ها نشان می‌دهد.

پیش‌بینی TOC توسط روش $\Delta \log R$

روش $\Delta \log R$ برای جدایش بین لاگ‌های تخلخل و مقاومت برای شناسایی و محاسبه TOC به کار برده می‌شود [۲]. رابطه زیر جدایش مابین لاگ‌های تخلخل و مقاومت ($\Delta \log R$) را نشان می‌دهد:



شکل ۹- نمایش گرافیکی و پلات‌های تصویری مقایسه‌ای TOC اندازه‌گیری شده برآورد شده، قبل و پس از خوشه‌بندی برای کل داده‌ها. عکس‌های تصویری مقدار کل کربن آلی برآورد شده: بعد از خوشه‌بندی (الف)، قبل از خوشه‌بندی (ب)، مقدار اندازه‌گیری شده (ج)، نرخ خطا برای هر دو حالت (د) و پلات در مقابل عمق برای همه حالت‌های برآورد شده و اندازه‌گیری شده (ه)



شکل ۱۰- نمایش مقدار کربن آلی اندازه‌گیری شده و برآورد شده با استفاده از شبکه عصبی و روش ΔLog در جاهای A, B و C

نتیجه‌گیری

از داده‌های ورودی و خروجی قبل از طراحی مدل‌های هوشمند می‌تواند به میزان قابل ملاحظه‌ای دقت تخمین‌های نهایی را بالا ببرد. در اینجا انطباق خوبی بین TOC‌های اندازه‌گیری شده حاصل از پیرولیز راک ایول و TOC‌های پیش‌بینی شده از کلاسترها وجود دارد (شکل ۶).

در مقایسه با روش $\Delta \log R$ ، کارایی روش‌های هوشمند به‌ویژه در صورت خوشه‌بندی بازه‌ها، بسیار بالاتر است. با استفاده از مدل شبکه عصبی مبتنی بر خوشه‌بندی، مقدار TOC برای چاه‌های A، B و C در میدان گازی پارس جنوبی تعیین شد. مقایسه بین TOC‌های اندازه‌گیری شده از پیرولیز راک ایول و مقدار پیش‌بینی شده از خوشه‌بندی براساس مدل شبکه عصبی و روش $\Delta \log R$ در شکل ۹ به نمایش درآمده است. همان‌طور که نشان داده شده، انطباق خوبی بین TOC‌های اندازه‌گیری شده و پیش‌بینی شده حاصل از روش شبکه عصبی در کل بازه مشاهده می‌گردد.

در این مقاله، از یک روش دو مرحله‌ای برای پیش‌بینی TOC از داده‌های پتروفیزیکی استفاده گردید. SOM و HCA برای تعیین الکتروفاسیس از داده‌های لاگ به‌کار گرفته شدند. آزمون اعتبارسنجی خوشه‌ها براساس مقدار Silhouette نشان داد که خوشه‌بندی به روش شبکه عصبی براساس روش SOM، نتایج مطلوب‌تری را برای دسته‌بندی داده‌ها و تعیین الکتروفاسیس ارائه می‌دهد. با روش الگوریتم SOM، ۷ کلاستر براساس داده‌های لاگ تعیین شد. ۷ مدل شبکه عصبی منطبق بر گروه‌های EF شناسایی شده، تولید گردید. پیش‌بینی در بازه‌های خوشه‌بندی شده مناسب بود و MSE اندازه‌گیری شده برابر ۰/۰۰۴۸ گزارش شد که در مقایسه با روشی که برای کل بازه عمقی طراحی شده بود و خطای ۰/۰۰۷۳ را نشان داد، دقت بهتری داشت. بنابراین می‌توان نتیجه گرفت که یک دسته‌بندی اولیه

مراجع

- [۱]. نجاری س.، علیزاده ب.، کدخدایی علی.، "تخمین کل کربن آلی (TOC) توسط روش‌های شبکه عصبی مصنوعی و $\Delta \log R$ در میدان گازی پارس جنوبی"، بیست و نهمین گردهمایی علوم زمین، سازمان زمین شناسی و اکتشافات معدنی کشور، تهران، ایران، ۱۳۸۹.
- [2]. Passey O. R., Moretti F. U. and Stroud J. D., "A practical modal for organic richness from porosity and resistivity logs", American Association of Petroleum Geologists Bulletin 74, pp.1777-1794. 1990
- [3]. Kamali M. R. and Mirshady A. A., Total organic carbon content determined from well logs using $\Delta \log R$ and neuro-fuzzy techniques. J. Petrol. Sci. Eng. 45, pp. 141-148. 2004.
- [4]. Huang Z., Williamson M. A., "Artificial neural network modeling as an aid to source rock characterization", Marine and Petroleum Geology 13 (2), pp. 227-290. 1996
- [5]. Kadkhodaie-Ilkhchi A., Rahimpour-Bonab H. and Rezaee M. R., "A Committee Machine with Intelligent Systems for Estimation of Total Organic Carbon Content from Petrophysical Data: An Example from the Kangan and Dalan Reservoirs in South Pars Gas Field", Iran. Computers & Geosciences 35, 459-474, 2009
- [6]. Kashfi M. S., "Greater Persian Gulf Permian-Triassic stratigraphic nomenclature requires study", Oil and Gas Journal (Tulsa) 6, pp. 36-44, 2000
- [7]. Ji C. Y., "Land-use classification of remotely sensed data using Kohonen self-organizing feature map neural networks", Photogrammetric Engineering and Remote Sensing 66, pp. 1451-1460. 2000,
- [8]. Vilmann T., Merenyi E. and Hammer B., "Neural maps in remote sensing image analysis", Neural Networks 16, pp. 389-403. 2003.

- [9]. Fayos J. and Fayos C., "*Wind data mining by Kohonen neural networks*", PLoS ONE 2, pp. 210. 2007.
- [10]. Cassano E. N., Lynch A. H., Cassano J. J. and Koslow M. R., "Classification of synoptic patterns in the western Arctic associated with extreme events at Barrow", Alaska, USA. *Climate Research* 30, pp. 83-97. 2006.
- [11]. Cole ou T., Poupon M. and Azbe K., "*Unsupervised seismic facies classification: a review and comparison of techniques and implementation*", *The Leading Edge* 22, pp. 942–953, 2003.
- [12]. Strecker U. and Uden R., "*Data mining of 3D poststack seismic attribute volumes using Kohonen self-organizing maps*", *The Leading Edge* 21, pp. 1032-1037. 2002.
- [13]. Kohonen T., Kaski S. and Lappalainen H., "Self-organized formation of various invariantfeature filters in the adaptive-subspace SOM", *Neural Computation* 9, pp. 1321-1344. 1997.
- [14]. Mukherjee A., "*Self-organizing neural network for identification of natural modes*", *The Journal of Computing in Civil Engineering* 11 (1), pp. 74-77. 1997.
- [15]. Astela A., Tsakovski S., Barbieri P. and Simeonov V., "*Comparison of self-organizing maps classification approach with cluster and principal components analysis for large environmental data sets*", *Journal of Water Research* 41, pp. 4566-4578, 2007.
- [16]. Kohonen T., "*Self-Organizing Maps, Springer series in Information Sciences*", New York, Springer-Verlag, Vol. 30, pp. 501, 2001.
- [17]. Tan P-N., "*Steinbach M. and Kumar V. Introduction to Data Mining*", Pearson Addison Wesley, pp. 769, 2006.
- [18]. Bhatt A., Helle H. B., "*Committee neural networks for porosity and permeability prediction from well logs*", *Geophysical Prospecting* 50, pp. 645-660. 2002.
- [19]. Bishop C. M., "*Neural Networks for Pattern Recognition*", Clarendon Press, Oxford, pp. 670, 1995.