Archive of SID

برآورد نگار ژئو شیمیایی با استفاده از الگوریتم ژنتیک مبتنی بر آنالیز خوشه‌بندی

زیبا حسینی[۱]، علی کدخدائی[۲]، سید محمداحسان طباطبائی[۳]

گروه زمین‌شناسی،دانشکده علوم طبیعی، دانشگاه تبریز

hosseini@sadi.ut.ac.ir

چکیده

نگارهای ژئوشیمیایی از جمله نگار مقدارکل کربن آلی (TOC)، کارآمدترین عامل ارزیابی سازندهای مستعد تولید هیدروکربن هستند، اما بدست آوردن این اطلاعات مستلزم صرف زمان و هزینه زیادی است. از این جهت، این پژوهش به برآورد نگار ژئوشیمیاییِ TOC با استفاده از نگارهای پتروفیزیکی که امروزه از تمامی چاه‌های حفاری تهیه شده و کم هزینه‌تر هستند، پرداخته و برای این منظور از الگوریتم خوشه بندی K-means و الگوریتم ژنتیک، بهره گرفته است. برای تخمین داده‌های TOC به این روش، کدی عددی با رویکردی دو مرحله ای طراحی شد. به طوری که در مرحله اول الگوریتم K-means سه خوشه بندی با تعداد خوشه‌های متفاوت تولید نموده و در مرحله‌ی دوم الگوریتم ژنتیکضرائب وزنی و نمائی از تابع غیرخطی پیشنهادی را برای داده‌های پتروفیزیکی در هر خوشه محاسبه می‌کند. سپس برای هر یک از خوشهبندی ها به تخمین داده های TOC پرداخته و با محاسبه مقدار میانگین مربع خطای کل در آنها، بهینه‌ترین خوشه بندی انتخاب می‌شود. در نهایت با استفاده از این مدل بهینه‌ی انتخاب شده، نگار TOCبرای کل داده‌ها محاسبه می‌گردد. این مطالعه بر روی داده‌های مربوط به سه چاهانجام گرفته است و قابل تعمیم به سایر چاه‌های اکتشافی خواهد بود.

واژه‌های کلیدی: مقدارکل کربن آلی(TOC)،نگارهای پتروفیزیکی، الگوریتم K-means، الگوریتم ژنتیک.

---

[۱] دانشجوی کارشناسی ارشد ژئوشیمی، دانشگاه تبریز

[۲] استادیار زمین شناسی نفت، دانشگاه تبریز

[۳] کارشناس ارشد مهندسی مکانیک-تبدیل انرژی، دانشگاه تهران

www.SID.ir

Archive of SID

Geochemical logs estimation using Genetic Algorithm based on clustering analysis

Hosseini, Z.[1], Kadkhodaie, A.[2], Tabatabaei, M.E.[3]

[1]*Department of Geology, Faculty of Natural Science, Tabriz University*
hosseini@sadi.ut.ac.ir

Abstract

Geochemical logs, such as Total Organic Carbon (TOC) are the effective criteria for evaluation the hydrocarbon production potential in formations. But gathering these information are oftentime consuming and expensive so ,in this study TOC geochemical logs were estimated by using petrophysical logs thatnowadays they are prepared from all wells drilled with low costs. For this purpose, K-means clustering and genetic algorithm were used. For estimation of TOC logs with this method, a numerical code was designed using a two-step procedure. In first step, the K-means algorithm generates three models with different number of clusters, and in the second step, genetic algorithm calculates weight and power coefficients of nonlinear function for petrophysical data in each cluster. Afterwards, the TOC logs were estimated and the best clustering model was chosen by calculating the mean squared error for all data. Finally, TOC logs could be determined for all input data by using this selected clustering model.This study was done based on data from three wells, and generalized to other exploration wells.

Keywords: Total Organic Carbon(TOC), Petrophysical log, K-means algorithm, Genetic Algorithm.

---

[1]M.SC. student in Geochemistry, Tabriz University
[2] Ph.D. in petroleum geology, Tabriz University
[3] M.SC. in Mechanical Engineering-Energy Conversion, University of Tehran

www.SID.ir

Archive of SID

۱- مقدمه

تحلیل نگار ژئوشیمیاییِ مقدار کل کربن آلی(TOC)، یکی از مهمترین عوامل در ارزیابی پتانسیل تولید و بررسی خصوصیات ژئوشیمیایی واحدهای هیدروکربنی می‌باشد[۱]. اهمیت داده‌های این نگار تا حدی است که امروزه اکثر شرکت‌های نفتی تهیه‌ی آن را برای اکتشاف نفت و گاز الزامی دانسته و بدون داشتن چنین اطلاعات ژئوشیمیایی اقدام به حفاری نمی‌کنند. گسترده‌ترین روش موجود برای تهیه نگارTOC، پیرولیز Rock-Eval می‌باشد که به دلیل هزینه‌ی بالای آن و وجود ناپیوستگی در برداشت مغزه‌ها، تنها معدودی از نمونه‌ها با این روش آنالیز می‌شوند. در این راستا، توسعه یک روش سریع و کم‌هزینه برای تخمین داده ژئوشیمیایی با صحتی قابل‌قبول اهمیتی منحصربه‌فرد دارد. برای رسیدن به این هدف، بسیاری از محققین سعی در برقراری ارتباط بین داده‌های TOC با داده‌های حاصل از نگارهای پتروفیزیکی که برای بیشتر چاه‌های حفاری قابل دسترس می‌باشند، نموده‌اند.از این جمله می‌توان بهPasseyو همکاران(۱۹۹۰)[۲]،Huangو Williamson(۱۹۹۶)[۳]، کمالی و میرشادی(۲۰۰۴)[۴]، کدخدایی و همکاران(۲۰۰۹)[۵]، سفیداری و همکاران(۲۰۱۲)[۶] اشاره نمود.این پژوهش ها نشانگر افزایش دقت با بکارگیری روش های تلفیقی و مدرن تر می باشند. بنابراین، پژوهش حاضربه معرفی رویکردی دو مرحله‌ای و مدرن جهت تولیدTOCمی‌پردازد. به طوری که از خوشه‌بندی مبتنی بر الگوریتمK-means استفاده نموده و پس از تولید خوشه‌ها با کمک الگوریتم ژنتیک، نگار TOC هر خوشه را برآورد می‌نماید. به این جهت، از نگارهای پتروفیزیکی و نتایج آنالیز ژئوشیمیایی خرده‌های حفاری سازندهای پابده، گورپی، کژدمی مربوط به سه چاه اکتشافی بهره گرفته شد.

۲- روش

۲-۱- الگوریتم خوشه‌بندیK-means

آنالیز خوشه‌بندی (Clustering Analysis) از جمله‌ی کارآمدترین روش‌ها در خوشه‌بندی و آنالیز آماری داده‌ها است. هدف آنالیز خوشه‌ای اندازه‌گیری شباهت و عدم شباهت بین داده‌ها و قرار دادن آنها در گروه‌هایی است که از لحاظ داخلی همگن و از لحاظ خارجی ناهمگن هستند[۸]. اندازه‌گیری میزان این شباهت بستگی به نوع کاربرد آن دارد. آنالیز خوشه‌ای به روش‌های مختلفی صورت می‌گیرد، از جمله آن، خوشه‌بندی به روش الگوریتم K-means است[۹]، که در این پژوهش از آن استفاده شد. این نوع خوشه‌بندی یک ابزار مفید و مؤثر برای زون‌بندی مخازن نفت و گاز محسوب می‌شود. در این الگوریتم نمونه‌ها بر اساس کمترین فاصله (معمولاً فاصله اقلیدسی) با مرکز خوشه‌هادر یکی از خوشه‌ها قرار می‌گیرند. به این ترتیب Kخوشه تشکیل خواهد شد که هر یک حاوی تعدادی نمونه است(تعداد خوشه‌ها K از قبل تعیین می‌شود).

۲-۲- الگوریتم ژنتیک

الگوریتمژنتیکدرواقعشناخته شدهترینوعازالگوریتم‌هایتکاملیاستکهطیسال‌های ۱۹۶۰ تا ۱۹۸۰ توسطجانهلند[1] وهمکارانشابداعوگسترشیافت[۱۰].اینالگوریتمدارایروندتکراریبودهکه جستجوراباجمعیتیازراهحل‌هایاولیهتصادفیآغازمی‌کند. اگرمعیار-هاینهاییارضانشود، دوعملگرخیلی مهم و متفاوت جهش[2]وادغام[3]در تولیدنسل بعدیبهکارگرفتهمی‌شوندتاجمعیتبهروزشود.درواقع، الگوریتمژنتیکفضایحلراباتکرارسهگامسادههجستجومی‌کند، که طرح کلی آن به این صورت است:

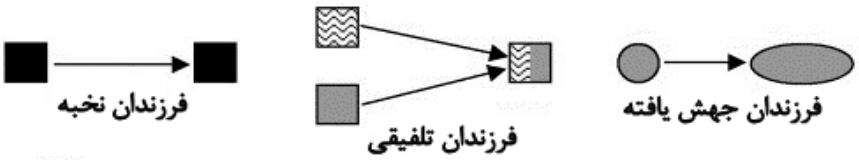
---

[1] *John Holland*
[2] *crossover*
[3] *mutation*

www.SID.ir

Archive of SID

- الگوریتم با تولید یک جمعیت آغازین تصادفی کار خود را آغاز می‌کند.
- سپس یک ترتیب از جمعیت جدید را ایجاد می‌نماید(نسل‌ها). این الگوریتم از سه روش برای تولید افراد نسل بعدی استفاده می‌نماید[۱۱] که به صورت شماتیک نشان داده شده است(شکل۱).



شکل۱) روش های تولید فرزند برای نسل بعدی از الگوریتم ژنتیک.

- الگوریتم زمانی متوقف می‌شود که یکی از شرایط توقف(تعداد نسل‌ها، محدودیت زمانی، محدودیت شایستگی و...) محقق شود.

۳- اجرای روش جهت تخمین TOC

در این پژوهش از الگوریتمK-means والگوریتم ژنتیک برای برآورد محتوی کربن آلی، از چهار لاگ پتروفیزیکی شامل مقاومت، نوترون، گاما و صوتی بهره گرفته شده است. مجموعه داده‌ها شامل ۱۴۳ داده‌ی آموزشی جهت ساخت مدل و ۶۰ داده‌ی تست جهت صحت‌سنجی آن می‌باشد، که این مجموعه داده‌ها از سه چاهِ یکی از در میادین نفتی تهیه شده است. برای ارزیابی و مقایسه بهتر، کلیه‌ی داده‌های ورودی و خروجی نرمالایز شدند.

برای بکارگیری این دو الگوریتم کدی عددی با رویکردی دو مرحله‌ای طراحیشد، که در مرحله اول الگوریتم K-meansبا ورودی هایی شامل داده‌های پتروفیزیکی و داده‌های عمق، خوشه ها را بر حسب K تعیین شده ایجاد می‌نماید. مقدار Kبه گونه‌ای تعریف شده است که از یک خوشه با گام دو تا ایجاد پنجخوشه پیش می‌رود و در نهایت سه خوشه‌بندی مجزا ارائه می‌دهد. در مرحله دوم الگوریتم ژنتیک، برای تخمین مقدار کل کربن آلی هرخوشه با استفاده از داده‌های پتروفیزیکی از معادله‌ی زیر استفاده می‌نماید:

$$TOC_{predicted}=C_1*GR^{C_2}+C_3*DT^{C_4}+C_5*LLD^{C_6}+C_7*CNL^{C_8}+C_9 \quad (۱)$$

C_1، C_2، C_3، C_4، C_5، C_6، C_7، C_8 به‌ترتیب مقادیر ضرایب وزنی و نمایی برای ورودی‌های GR[۱]، DT[۲]، LLD[۳]،CNL[۴] هستند و C_9مقدار ثابت معادله است.

تابع هدف نیز برای هر خوشه طبق فرمول زیر مینیمم می‌گردد:

$$MSE=\sum_{i=1}^{N} 1/N \left(TOC_{real}-TOC_{predicted}\right)^2 \quad (۲)$$

در این معادله MSE میانگین مربع خطا،TOC_{real} مقدار کل کربن آلی اندازه‌گیری‌شده توسط پیرولیز Rock-Eval، $TOC_{predicted}$ مقدار کل کربن آلی پیش‌بینی شده توسط الگوریتم ژنتیک وN تعداد نمونه‌ها می‌باشند.

---

[1] *Gamma-Ray*
[2] *Transit time (ΔT)*
[3] *DLL (DaulLaterolog)*
[4] *Compensated neutron log*

www.SID.ir

Archive of SID

برای الگوریتم ژنتیک، سایز جمعیت آغازی ۴۰ در نظر گرفته شد. این مشخصه نشان دهنده‌ی تعداد افراد در هر نسل بوده که طی فرآیندی تصادفی ایجاد می‌گردند. حداکثر نسل ۲۰۰۰ و محدودیت واماندگی نسل‌ها ۵۰۰ انتخاب گردید. محدودیت واماندگی نسلی یکی از معیارهای توقف است، بدین معنی که اگر در طول ۵۰۰ نسل هیچ گونه پیشرفتی حاصل نگردد الگوریتم متوقف شود. تعداد فرزندان نخبه برای هر نسل ۴ و کسر فرزندان تلفیقی ۰/۷۵(تعداد ۲۷ فرزند) تنظیم شد، که ۹ فرد باقیمانده از نسل بعدی توسط عملیات جهش تولید می‌گردند. دقت این الگوریتم برای تابع هدف نیز، 10^{-26} تعریف شد.

پس از اجرای برنامه برای داده‌های آموزشی، الگوریتم ژنتیک بهترین ضرایب وزنی و نمایی از معادله(۱) را برای هر خوشه برآورد - می‌کند. سپس این معادله ها بر روی داده‌های تست اعمال شده و نگار TOC را پیش‌بینی می‌نمایند.در انتها نتایج خوشه‌ها با هم ترکیب شده و مقدار میانگین مربع خطا کل(MSE_t)طبق معادله (۳) و همچنین ضریب همبستگی(R^2) برای هر خوشه‌بندی به عنوان معیاری از کارایی مدل، محاسبه می‌گردد(جدول۱).

$$MSE_t = \frac{\sum_{i=1}^{k} MSE_i \times n_i}{\sum_{i=1}^{k} n_i} \quad (۳)$$

در این معادله، MSE_i مقدار میانگین مربع خطا مربوط به هر خوشه، n_i تعداد نمونه در هر خوشه و k تعداد کل خوشه‌ها می‌باشد.

تعداد خوشه‌های زیاد گاهی به دلیل کمبود داده‌های TOC ممکن است خطاهای بزرگی در پیش بینی داده‌های تست ایجاد نماید، اما نتایج این پژوهش نشان می‌دهد که افزایش تعداد خوشه تا حدی معقول، تأثیر خوبی در بهبود جواب خواهد گذاشت. لازم به ذکر است که این پژوهش از داده‌های سه سازند استفاده نموده و اگرچه داده‌های پتروفیزیکی موجود انعکاسی از عمق و دیگر خصوصیات هریک از سازندها را ارائه می‌دهند اما در روند این پژوهش به این نتیجه دست یافتیم که تأثیر مستقیم داده‌های عمق باعث ایجاد خوشه‌های بهینه‌تری می‌گردد. همانطور که جدول (۱) نشان می‌دهد مدلی با پنج خوشه مدل بهینه شناخته شده است.MSE_tبرای داده‌های آموزشی این مدل ۰/۰۰۰۴۱۸ و برای داده‌های تست آن ۰/۰۰۱۲۰۷ بدست آمد.مقدار ضریب همبستگی(R^2) بین TOC واقعی و TOC حاصل از الگوریتم ژنتیک برای داده‌های آموزشی و تستِ این مدل همدر شکل (۲) آمده است.همچنین در شکل(۳)، تطابق بین TOC اندازه‌گیری شده و TOC پیش‌بینیشده در کل داده‌های این مدل به‌همراه مقدار خطا نشانداده‌شده‌هاست. (شکل۴) نیز به گونه‌ای پویاتر تطبیق گرافیکی بین TOCواقعی نرمالایز شده وTOCحاصل از الگوریتم ژنتیک با ذکر نام خوشه در داده‌های تست را به تصویر کشیده است.

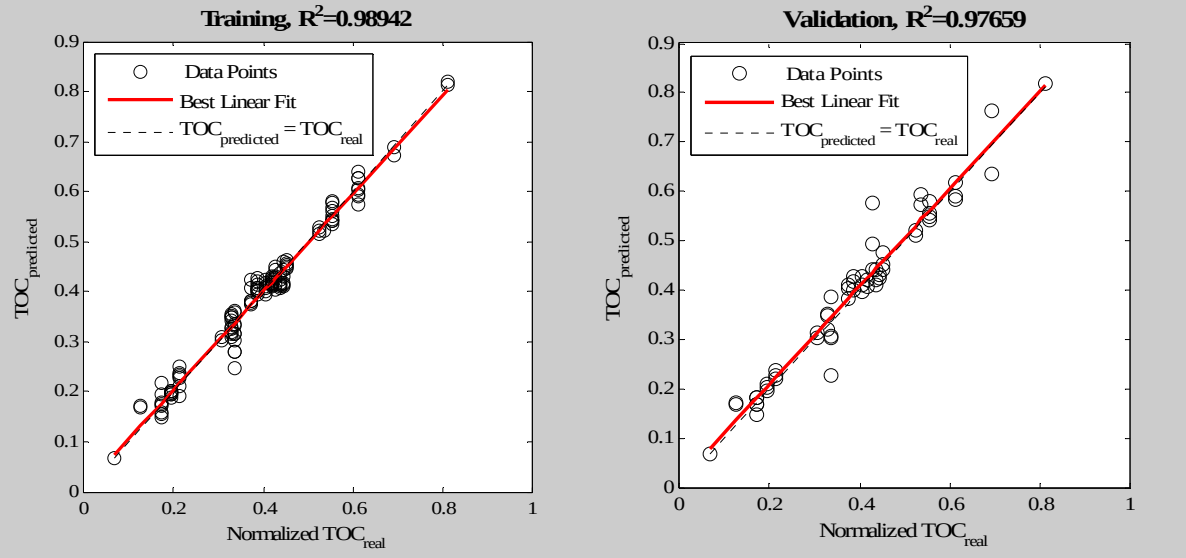
جدول۱)مقادیرMSE_t و R^2حاصل از الگوریتم ژنتیک برای هر خوشه‌بندی.

تعداد خوشه(K)	R^2 در داده‌های تست	MSE_tدر داده‌های تست
*۱	۰/۹۳۸۹۲	۰/۰۰۲۹۲۷
۳	۰/۹۵۳۱۳	۰/۰۰۱۹۶۰۱
۵	۰/۹۷۶۵۹	۰/۰۰۱۲۰۷

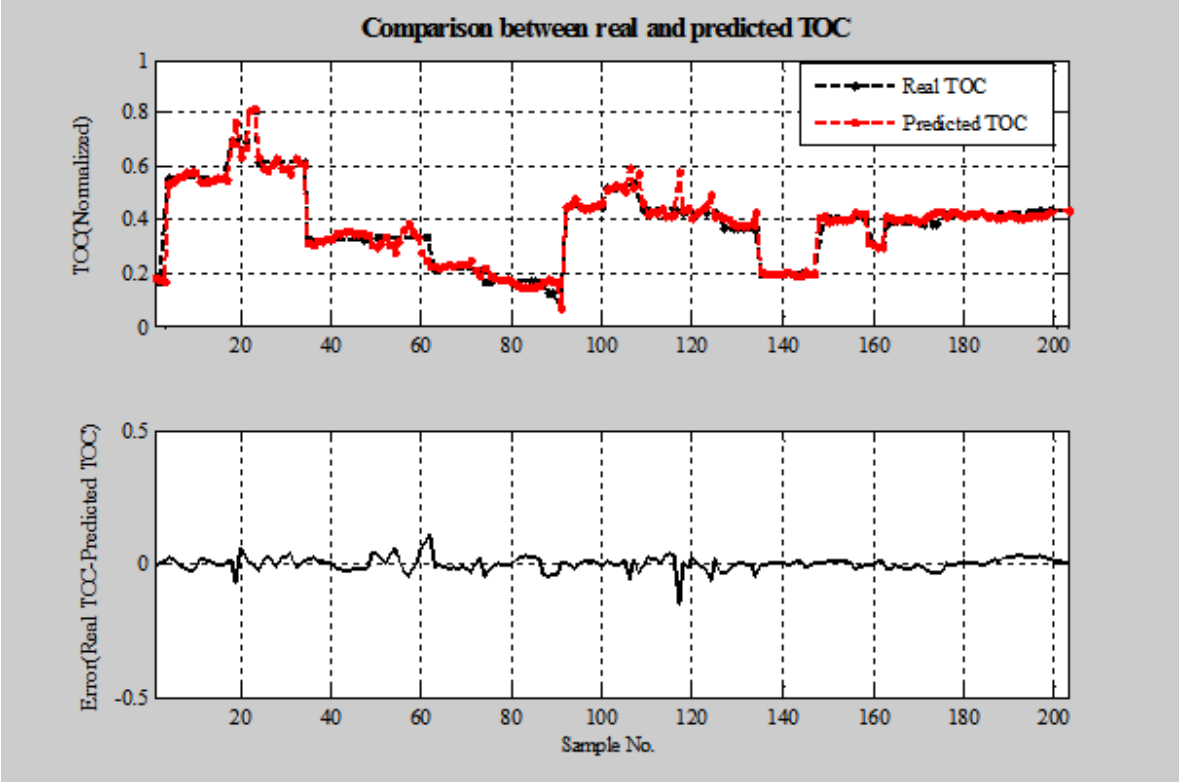
*برابر با قبل از خوشه بندیK-meansمی باشد.

www.SID.ir

Archive of SID



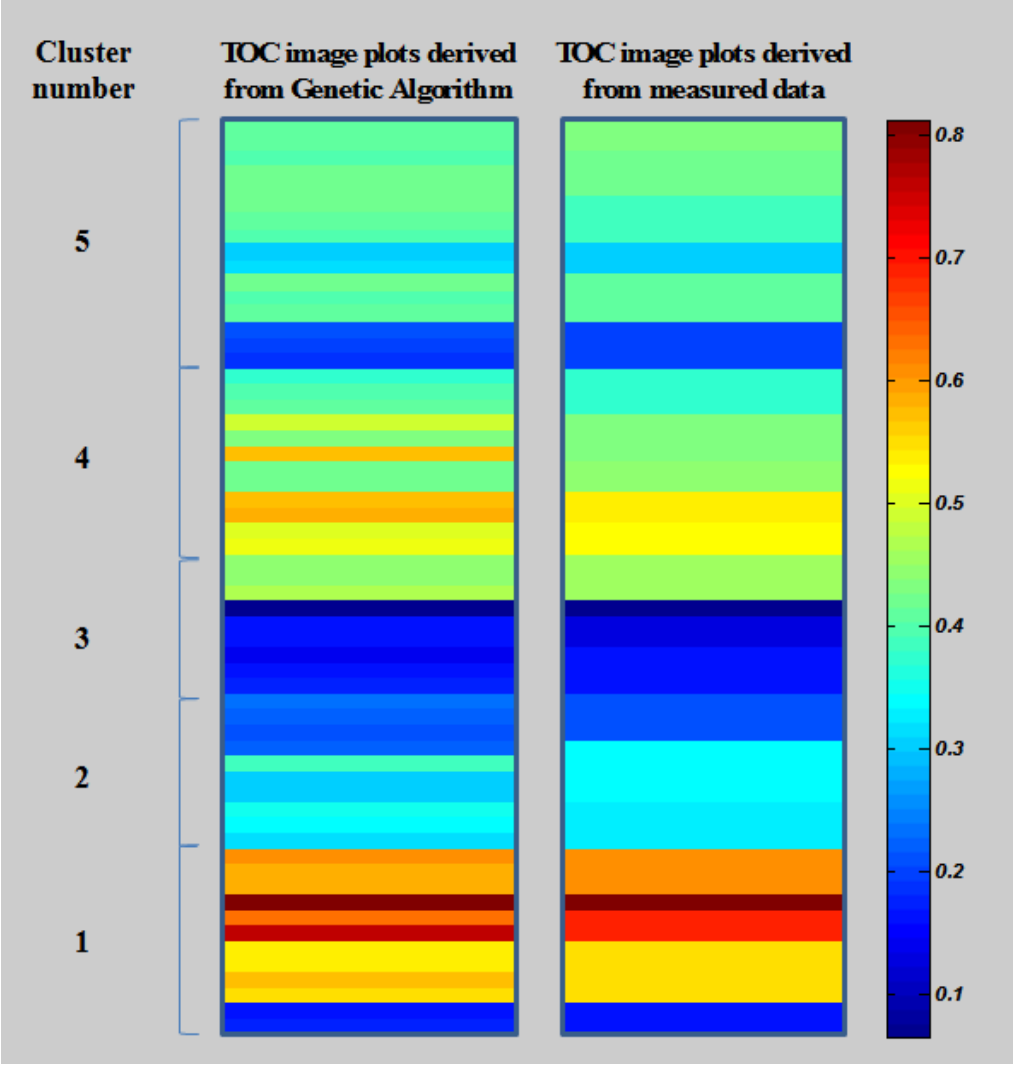
شکل۲) ضریب همبستگی بین TOC برآورد شده و TOC واقعی در داده‌های آموزشی و داده های تست.



شکل۳) نمایش تطابق نموداری بین TOC واقعی نرمال و TOCبرآورد شده در کل داده‌ها.

www.SID.ir

Archive of SID



شکل۴) نمایش تطابق گرافیکیبین TOC واقعی نرمال و TOCبرآورد شده در هر خوشه ازداده‌های تست.

۴- نتیجه گیری

در این مطالعه، از الگوریتم خوشه‌بندی K-means و الگوریتم ژنتیک جهت تخمین داده‌های TOC از داده‌های پتروفیزیکی استفاده شد. داده‌های پتروفیزیکی (مقدار مقاومت، زمان عبور صوت، نوترون و همچنین گاما) به‌عنوان ورودی و داده‌های TOC حاصل از آنالیز ژئوشیمیایی به‌عنوان خروجی مطلوب شبکه تعیین گردید. نتیجه‌ی صحت‌سنجی شبیه‌سازی با این روش نشان داد که افزایش خوشه‌بندی در حد معقول و در صورتی که با کمبود داده مواجه نباشیم می‌تواند در بهبود برآوردها مؤثر باشد. به همین دلیل مدلی با پنج خوشه مدل بهینه شناخته شد. مقادیر MSE و R^2برای داده‌های تست این مدل به‌ترتیب ۰/۰۰۱۲۰۷و ۰/۹۷۶۵۹ارائه گردید.مدل ارائه شده در این پژوهش به صورت یک مطالعه موردی بر روی داده‌های مربوط به سه چاه انجام گرفته است وقابل تعمیم به برآورد نگار ژئوشیمیایی سایر چاه‌ها می‌باشد. به امید آن‌که این پژوهش گامی کوچک در پیشرفت صنعت ملی نفت و بهره‌وری از ذخایر و پتانسیل‌های هیدروکربنی باشد.

www.SID.ir

Archive of SID

مراجع

[1] Alizadeh, B., Najjari, S., Kadkhodaie-Ilkhchi, A. 2011. Artificial neural network modeling and cluster analysis for organic facies and burial history estimation using well log data: A case study of the South Pars Gas Field, Persian Gulf, Iran. Computers &Geosciences, 45: 261–269.

[2] Passey, O.R., Moretti, F.U., Stroud, J.D., 1990. A practical modal for organic richness from porosity and resistivity logs.American Association of Petroleum Geologists Bulletin, 7: 1777–1794.

[3] Huang, Z., Williamson, M.A., 1996. Artificial neural network modeling as an aid to source rock characterization. Marine and Petroleum Geology, 13 (2): 227–290.

[4] Kamali, M.R., Mirshady, A.A., 2004. Total organic carbon content determined from well logs using Δ log R and neuro-fuzzy techniques. Journal of Petroleum Science and Engineering, 45: 141–148.

[5] Kadkhodaie-Ilkhchi A, Rahimpour-Bonab H, Rezaee M.R. 2009. A committee machine with intelligent systems for estimation of total organic carbon content from petrophysical data: An example from Kangan and Dalan reservoirs in South Pars Gas Field, Iran. Computers & Geoscience, 35(3):459-474.

[6] Sefidari, E., Kadkhodaie-Ilkhchi, A., Najjari, S., 2012. Comparison of intelligent and statistical clustering approaches to predicting total organic carbon using intelligent systems.Journalof PetroleumScience and Engineering , 86-87: 190–205.

[7] Menzies, T., Pecheur, C. 2005. Verification and Validation and Artificial Intelligence. Advances in Computers,65: 154-203.

[8] MATLAB User's Guide, 2011. Version 7.12, Statistics Toolbox.The MathWorks Inc.

[9] Han, J., Kamber, M., " Data Mining: Concepts and Techniques", Second Edition, Morgan Kaufmann Publishers, 2006.

[10] John H. Holland, 1975. Adaptation in Natural and Artificial Systems. University of Michigan Press (Ann Arbor).

[11]Maschio, C., Campane Vidal, A., Schiozer, D.J. 2008, A framework to integrate history matching and geostatistical modeling using genetic algorithm and direct search methods, Journal of Petroleum Science and Engineering, 86: 34-42.

www.SID.ir