

مدل‌سازی تخلخل توسط رویکرد تئوری بیزین ترکیب اطلاعات و مقایسه آن با شبکه عصبی چندلایه و رگرسیون خطی چندگانه در میدان نفتی آزادگان

عطیه مظاهری طرئی*^۱، حسین معماریان^۲، بهزاد تخم‌چی^۳، بهزاد مشیری^۴

۱- دانش آموخته اکتشاف نفت، دانشگاه تهران، atie_mazaheri@ut.ac.ir

۲- عضو هیئت علمی دانشگاه تهران

۳- عضو هیئت علمی دانشگاه صنعتی شاهرود

۴- عضو هیئت علمی دانشگاه تهران

چکیده

پارامتر تخلخل یکی از مهم‌ترین خصوصیات مخزن می‌باشد که با مطالعه مغزه به‌دست می‌آید. با این وجود تمامی چاه‌های یک میدان دارای مغزه نیستند. هم‌چنین در برخی از چاه‌ها مانند چاه‌های افقی مغزه‌گیری عملاً غیر ممکن است. ولی تقریباً در تمامی چاه‌ها نمودارگیری صورت می‌گیرد. به‌طور معمول از نمودارهای چاه‌نگاری به‌منظور تخمین تخلخل نیز استفاده می‌شود. تخلخلی که از این نگارها به‌دست می‌آید تحت تأثیر عواملی هم‌چون دما، فشار، نوع سیال، میزان هیدروکربور و شیل موجود در سازند قرار می‌گیرند و در نتیجه با میزان واقعی تخلخل کمی متفاوت است. بنابراین تخمین‌ها توأم با خطا و عدم قطعیت هستند. شاید بهترین و در عین حال عملی‌ترین روش جهت کاهش عدم قطعیت تخمین، استفاده از منابع مختلف داده جهت تخمین و در واقع استفاده از تکنیک‌های ترکیب اطلاعات باشد. کارکرد اصلی این تکنیک‌ها، افزایش اطمینان و کاهش ریسک در تصمیم‌گیری‌ها است. در این تحقیق، برای تعیین مقادیر تخلخل، با استفاده از داده‌های چهار چاه واقع در میدان نفتی آزادگان، ابتدا از دو تکنیک شبکه عصبی چندلایه و رگرسیون خطی چندگانه استفاده شده و در نهایت نتایج این روش‌ها با تکنیک ترکیب اطلاعات (تئوری بیزین) مقایسه شده است. برای بررسی قابلیت تعمیم این سه روش در هر تکنیک، پارامتر تخلخل در یک چاه دیگر میدان نیز تخمین زده شده است. تعداد متغیرهای ورودی برای تخمین تخلخل در چاه مورد مطالعه در روش‌های شبکه عصبی و رگرسیون خطی چندگانه ۷ است و در تکنیک ترکیب اطلاعات نیز از حداکثر ۷ متغیر ورودی استفاده شده است. در نهایت با مقایسه نتایج حاصل از ۳ روش نامبرده مشاهده شده است که تکنیک ترکیب اطلاعات (تئوری بیزین) از اعتبار بالاتری برخوردار است و در تخمین تخلخل به مقدار قابل توجهی از دو تکنیک شبکه عصبی چندلایه و رگرسیون خطی چندگانه بهتر عمل نموده است؛ به نحوی که همبستگی نتایج با واقعیت بیش از ۹۰٪ به‌دست آمده است.

کلمات کلیدی: عدم قطعیت، تخمین، ترکیب اطلاعات، تئوری بیزین، شبکه عصبی چندلایه، رگرسیون خطی چندگانه، میدان آزادگان، سازند سروک، ایران

۱. مقدمه

تخلخل یک سنگ را می‌توان به دو دسته اولیه و ثانویه تقسیم بندی کرد. تخلخل اولیه در حین رسوب‌گذاری و تخلخل ثانویه در اثر فرآیندهای پس از رسوب‌گذاری شکل می‌گیرد. تخلخل وابسته به اندازه دانه‌ها و نحوه قرارگیری دانه‌ها و همین‌طور به مقدار سیمان یا مواد پرکننده و عمق تدفین بستگی دارد. تخلخل یک سنگ مخزن توسط مطالعات مغزه، تکنیک‌های چاه‌پیمایی و آزمایش چاه، قابل تعیین است. ولی هنگامی که حجم فضاهای خالی یک سنگ اندک باشد، اندازه گیری تخلخل با روش‌های ذکر شده بسیار دشوار است. در این‌گونه موارد از روش تزریق جیوه استفاده می‌شود [۱].

تخلخل یکی از سه متغیر مهم در ارزیابی ویژگی‌های یک میدان نفتی است. اطلاعات مربوط به تعیین متغیرهای پتروفیزیکی، غالباً با استفاده از مطالعه مغزه‌ها، حاصل می‌شوند؛ ولی استفاده از این روش در بسیاری موارد قابل اجرا نیست. از روش‌ها و روابط تجربی نیز به دلیل داشتن مشکلات خاص آن‌ها و تعلق نتایج مربوط به آن‌ها به منطقه ای خاص، نمی‌توان با اطمینان کامل استفاده کرد [۲].

تاکنون روش‌های متنوعی برای تخمین تخلخل در یک مخزن نفتی معرفی شده‌اند، که دامنه‌ای از روش‌های سنتی تا روش‌های هوشمند را شامل می‌شود. در مقابل روش‌های سنتی، روش‌های هوشمند، که در اصطلاح محاسبه نرم^{۳۸} نامیده می‌شوند، قرار دارند. از جمله تکنیک‌های محاسبات نرم می‌توان به شبکه‌های عصبی مصنوعی و منطق فازی اشاره کرد. این روش‌ها نقش پررنگی در پردازش داده‌های پتروفیزیکی ایفا می‌کنند. ممکن است روش‌های سنتی در مخازن همگن پاسخگو باشند ولی در مخازن ناهمگن، اعتبار همبستگی‌ها کاهش می‌یابد [۳].

در سال‌های اخیر توجه روزافزونی بر روی مساله‌ی تلفیق اطلاعات تولید شده از منابع و سنجنده‌های مختلف به منظور افزایش قابلیت‌ها و توانایی‌های سیستم‌های هوشمند معطوف گردیده است [۴].

تکنیک‌های تلفیق اطلاعات، داده‌های اخذ شده توسط سنجنده‌های مختلف به همراه اطلاعات مربوطه را با هدف حصول به استنتاج بهتر و مشخص‌تر از آنچه که با بکارگیری هر یک از اطلاعات به شکل منفرد و مستقل بدست می‌آید، با یکدیگر ترکیب می‌کنند. از این‌رو موضوع تلفیق اطلاعات، یکی از موضوعات مورد علاقه پژوهش‌گران در چند سال اخیر بوده است [۵، ۶ و ۷]. با اینکه در تصمیم‌گیری بر مبنای اطلاعات هر منبع به صورت جداگانه، تصمیمات، بعضاً^۱ نتایج مختلفی را به دست می‌دهد، تصمیم‌گیری بر مبنای ترکیب اطلاعات نه تنها مشکل اختلاف در تصمیم‌گیری منابع گوناگون را ندارد، بلکه به علت استفاده از تمام منابع اطلاعاتی ممکن، اطمینان و قطعیت تصمیم‌گیری بیشتری را دارد. در این مقاله از روش نوین ترکیب اطلاعات (ثوری بی‌زین) جهت تخمین تخلخل و مقایسه با نتایج حاصل از تخمین به روش‌های رگرسیون خطی چندگانه و شبکه عصبی، استفاده شده است.

۲. تخلخل

تخلخل قسمتی از سنگ است که توسط دانه‌ها پر نشده و به صورت ریاضی با رابطه ۱ تعریف می‌شود [۸]:

$$\varphi = \frac{V_b - V_{gr}}{V_b} = \frac{V_p}{V_b} \quad (1)$$

که در آن φ = تخلخل، V_b = حجم کپه‌ای سنگ مخزن، V_{gr} = حجم دانه‌ها و V_p = حجم فضاهای خالی است.

^۱ Soft Computing

تخریل مخازن ماسه‌ای معمولاً کمتر از درجه تخریل مخازن آهکی است. اگر تخریل مخزنی کمتر از ۵٪ باشد، حفاری از نظر اقتصادی مقرون به صرفه نیست، چون نمی‌تواند نفت زیادی را در خود ذخیره نماید. در جدول ۱ کیفیت مخازن از نظر میزان تخریل طبقه‌بندی شده است.

جدول ۱. کیفیت مخازن از نظر میزان تخریل [۹].

کیفیت مخزن	درجه تخریل (%)
قابل اغماض	۰-۵٪
ضعیف	۵-۱۰٪
متوسط	۱۰-۱۵٪
خوب	۱۵-۲۰٪
خیلی خوب	۲۰-۲۵٪

۳. رگرسیون خطی چندگانه

با روش‌های مطالعاتی چند متغیره، می‌توان هم‌زمان به تحلیل و بررسی چندین متغیر مختلف پرداخت. برای دسترسی به نتایج مطلوب‌تر از این روش‌ها، نیازمند به نمونه‌های فراوان و در عین حال درست است؛ زیرا این روش‌ها در مقابل اطلاعات نادرست، حساسیت بالایی دارند و ورود چنین داده‌هایی ممکن است منجر به بروز خطاهای بزرگی در نتایج به دست آمده شود. افزون بر این، برای استفاده از این روش‌ها، متغیرها باید توزیع نرمال داشته باشند و تغییر آن‌ها از یک رابطه خطی پیروی کند.

رگرسیون خطی چند متغیره در حقیقت، ارتباط بین یک سری از متغیرهای مستقل را با یک متغیر مورد نظر بیان می‌کنند. در صورت وجود متغیرهای مستقل x_1, x_2, \dots, x_n اگر بخواهیم ارتباط خطی بین آن‌ها و متغیر y که وابسته به آن‌هاست ایجاد کنیم، رابطه ۲ باید بین آنها برقرار باشد:

$$y = a_0 + a_1x_1 + a_2x_2 + \dots + a_nx_n + e \quad (2)$$

که در این رابطه، از مقادیر a_1, a_2, \dots, a_n با عنوان ضرایب رگرسیون یاد می‌شوند. این ضرایب، ضرایب نامشخصی هستند که در حقیقت، مسئول برآورد متغیر وابسته‌اند. در صورتی که از طرفین رابطه فوق، امید ریاضی گرفته شود، به دلیل اینکه امید ریاضی مقدار خطای e برابر با صفر می‌باشد، می‌توان رابطه ۳ را نوشت:

$$E(y) = a_0 + a_1x_1 + a_2x_2 + \dots + a_nx_n \quad (3)$$

که $E(y)$ در حقیقت، مقدار مورد انتظار تابع تحت تاثیر و ورود مقادیر متغیرهای x_1, x_2, \dots, x_n است [۱۰].

۴. شبکه عصبی چندلایه

شبکه‌های عصبی مصنوعی یکی از مهم‌ترین روش‌های محاسبه نرم محسوب می‌شوند، که با پردازش داده‌های تجربی، دانش یا قانون نهفته در دل داده‌ها را به ساختار شبکه منتقل می‌کنند. این سیستم‌ها سعی در مدل‌سازی ساختار عصبی مغز بشر دارند [۱۱]. در سال‌های اخیر، شبکه‌های عصبی به‌عنوان ابزاری قدرتمند برای مدل‌سازی سیستم‌های پیچیده مطرح شده‌اند. با

استفاده از مجموعه‌ای از ورودی‌ها و اندازه‌گیری‌های هدف، این روش قادر به آموزش دیدن و استنتاج روابط غیرخطی پیچیده بین آن‌ها خواهد بود.

در یک فرایند عصبی پردازش داده‌ها، مجموعه داده‌ها به سه دسته آموزشی، اعتبارسنجی و آزمون تقسیم می‌شوند. دسته داده‌های آموزشی به منظور تنظیم وزن‌های شبکه به کار می‌روند. در این فرایند، خروجی ایده‌آل نیز در دسته داده‌های آموزشی به شبکه داده می‌شود. دسته داده‌های اعتبارسنجی برای اطمینان از یادگیری درست شبکه به کار می‌روند. پس از آنکه فرایند یادگیری با موفقیت به پایان رسید، داده‌های آزمون روی شبکه اعمال می‌شوند تا از یادگیری درست شبکه اطمینان حاصل گردد [۱۲].

۵. ترکیب اطلاعات

تلفیق اطلاعات ترکیب تجمعی اطلاعات به دست آمده از منابع متعدد دانش، نظیر سنجنده‌ها، به منظور فراهم ساختن درک صحیح و بهتر، از موضوع می‌باشد [۱۳]. تکنیک‌های ترکیب اطلاعات، اطلاعات منابع مختلف و سایر اطلاعات مرتبط را با هم ترکیب می‌کنند؛ به نحوی که بتوان استنباط‌های خاصی را ارائه کرد که پیش‌تر توسط یک منبع اطلاعاتی مستقل قابل ارائه نبوده‌اند [۱۴]. مهم‌ترین مزیت استفاده از ترکیب اطلاعات منابع چندگانه، این است که اطلاعات هر منبع اطلاعاتی یا عدم-قطعیت دارد و یا دقیق نیست. هم‌چنین معمولاً یک ویژگی نمی‌تواند به‌تنهایی هدف موردنظر را تامین کند و حتی یک منبع اطلاعاتی، معمولاً نمی‌تواند یک نمونه را از بقیه‌ی نمونه‌ها با دقت مورد انتظار تشخیص دهد [۱۵]. به طور کلی، هدف از پردازش داده‌های حاصل از منابع مختلف، تصمیم‌گیری در شرایط عدم قطعیت می‌باشد [۱۶]. شاید امروزه در دنیای علم، محدود نمودن تکنیک‌های ترکیب اطلاعات میسر نباشد؛ چرا که بسیاری اوقات پژوهش‌گران، اطلاعات منابع مختلف را ترکیب می‌کنند، بدون این که نام ترکیب اطلاعات را در گزارشاتشان بیاورند. از طرف دیگر، محققان ترکیب اطلاعات، سعی در کلاسیک نمودن و طبقه‌بندی مفاهیم ترکیب اطلاعات به عنوان یک علم کاربردی دارند.

از تکنیک‌های متداول در ترکیب اطلاعات می‌توان به متغیرهای آماری، تکنیک احتمالات شرطی بیزین، دمپسترشیفر، میانگین وزنی مرتبه یافته، فیلتر کلمن، روش‌های کرنل و غیره اشاره کرد. روش‌های هوشمند نسبت به روش‌های متداول دارای درصد موفقیت و سازگاری بیشتری با شرایط مختلف هستند، اما پیچیدگی مدل‌هایشان نیز بیشتر است. کاربردی‌ترین تکنیک‌های هوشمند در ترکیب اطلاعات عبارت‌اند از شبکه‌های عصبی مصنوعی و فازی [۱۷]. در این مقاله، از تئوری احتمالات شرطی بیزین، برای تخمین تخلخل در یکی از مخازن نفتی میدان آزادگان استفاده شده است.

۵-۱. مبانی تئوری احتمالات شرطی بیزین

توماس بیز برخی روش‌های اساسی را در تئوری احتمالات ایجاد کرد. تئوری بیزین بیانگر رابطه بین احتمالات شرطی در تئوری احتمال است. فرض مهم در به‌کارگیری آن، این است که دانش اولیه یا پیشین را به‌توان با یک توزیع احتمال مدل کرد [۱۸]. وقتی که یک پدیده خاص و دقیق دیده می‌شود، یک دانش ثانویه و یا پسین را می‌توان با به‌کارگیری توزیع احتمال شرطی به دست آورد. در واقع تئوری بیزین یک روش عددی برای به‌روز رسانی^{۳۹} احتمال یک فرضیه بنا بر یک مشاهده است. احتمال بیز از رابطه ۴ قابل محاسبه است.

$$P(W_i|X) = \frac{P(X \cap W_i)}{P(X)} = \frac{P(X|W_i) P(W_i)}{\sum_{j=1}^c P(X|W_j) P(W_j)} \quad (4)$$

^{۳۹}Update

در این رابطه X متغیر ورودی، Wi کلاس مورد نظر و $P(Wi \setminus X)$ احتمال پسین صحیح بودن کلاس Wi برای متغیر X می باشد [۱۹].

اگر چه به نظر می رسد که تئوری بیز برای پیشگویی مناسب می باشد، اما نکاتی وجود دارد که استفاده از این تئوری را در ترکیب اطلاعات محدود می کند. نقطه ضعف این تئوری آن است که عدم قطعیت منابع اطلاعاتی را در نظر نمی گیرد و با توجه به پیچیدگی در اندازه گیری احتمالات پسین، می تواند خطا داشته باشد [۱۸].

۶. معرفی میدان مورد مطالعه

میدان نفتی مورد مطالعه، یکی از بزرگترین میداین نفتی ایران است که در ۸۰ کیلومتری غرب اهواز و نزدیک مرز ایران و عراق، در پهنه آبادان و قسمت غربی فروافتادگی دزفول واقع است. در این میدان چهار سازند سروک، کزدمی، گدوان، فهلیان، سازند نفتی ایلام و لایه ماسه سنگی کزدمی به عنوان سازندهای مورد بهره برداری شناخته شده اند. چاههایی که در این مطالعه مورد بررسی واقع شدند، ۵ چاه مربوط به مخزن سروک واقع در این میدان می باشند. سازند سروک به ۱۲ بخش تقسیم بندی شده است.

بخش های مخزنی ($S_{11}, S_9, S_7, S_3, S_1$) دارای تخلخل متوسط تا عالی؛ که بیشتر از نوع حفره های هستند تا درون دانه ای؛ می باشند. بخش های غیر مخزنی (S_{12}, S_{10}, S_6, S_2) دارای تخلخل پایین تا متوسط و گاهاً ضعیف هستند؛ که بیشتر از نوع حفره های می باشد تا درون دانه ای؛ سازند حد متوسط (S_5) دارای تخلخل متوسط تا عالی و گاهاً کم می باشد که آن نیز بیشتر از نوع حفره های است. نواحی غیر مخزنی تا مخزنی متوسط سازند سروک (S_4, S_8) دارای تخلخل پایین تا متوسط و گاهاً خوب و از نوع حفره های می باشند. تخلخل از نوع قالبی در این دوازده ناحیه وجود ندارد. طول کلی مغزه گرفته شده از ۵ چاه از سازند سروک در این میدان ۳۴۹ متر می باشد که بیشتر قسمت های مغزه گیری شده (۳۳۰ متر) مربوط به قسمت های بالایی سازند سروک است [۲۰].

۷. تخمین تخلخل با استفاده از رگرسیون خطی چند گانه و شبکه عصبی چند لایه

به منظور شناسایی داده های ورودی، که شامل داده های ۴ چاه از میدان ذکر شده می باشد، ابتدا ضرایب همبستگی بین چاه- نمودارهای مختلف با تخلخل تعیین شد. سپس، داده هایی که بیشترین مقدار مطلق همبستگی را با تخلخل داشتند، به عنوان داده های ورودی انتخاب شدند. به این منظور، ضرایب همبستگی برای چاه نمودارهای مختلف و تخلخل توسط نرم افزار SPSS به روش اسپیرمن محاسبه و نتایج آن در جدول ۲ درج شده است.

جدول ۲. ضریب همبستگی بین تخلخل و سایر متغیرها در چاه مورد مطالعه.

تخلخل	CGR	DT	LLD	LLS	MSFL	NPHI	RHOB	
۱	-۰/۲۴۶	۰/۸۴۸	۰/۳۷۰	۰/۲۸۰	۰/۰۱۲	۰/۷۷۱	-۰/۸۵۵	ضریب همبستگی با تخلخل

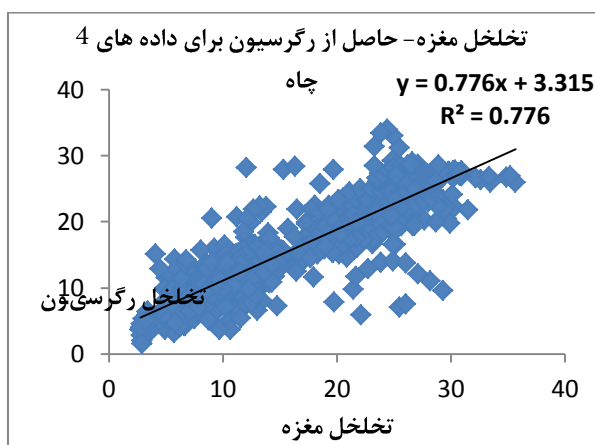
طبق جدول ۲، متغیرهای ورودی مورد استفاده در این تحقیق، شامل چاه نمودارهای صوتی (DT)، نوترون (NPHI)، چگالی (RHOB)، مقاومت ویژه لئولوگ عمیق (LLD)، مقاومت ویژه لئولوگ کم عمق (LLS)، پرتو گاما (CGR) و مقاومت ویژه کروی میکرو (MSFL) می باشد. در نتیجه از ۷ متغیر ورودی جهت تخمین تخلخل به روش رگرسیون چند متغیره و شبکه عصبی استفاده شده است.

۱-۷. استفاده از رگرسیون خطی چندگانه

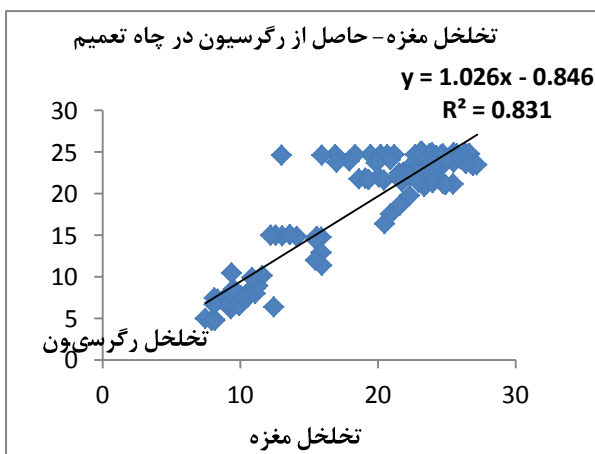
پس از انتخاب متغیرهای ورودی رگرسیون خطی چندگانه، به منظور تخمین تخلخل، از داده‌های چاه‌نگارها و مغزه ۴ چاه از میدان مورد مطالعه استفاده شد. پس از تعیین وزن‌های مختلف مربوط به متغیرهای مورد استفاده در معادله رگرسیون، معادله رگرسیون تخلخل به صورت معادله ۵ به دست آمد:

$$\text{Porosity} = -41.8982 * \text{RHOB} + 3.8287 * \text{NPHI} + 0.1430 * \text{MSFL} - 0.0395 * \text{LLS} + 0.0106 * \text{LLD} + 0.0794 * \text{DT} - 0.0183 * \text{CGR} + 99.9912 \quad (5)$$

نتیجه حاصل از تخمین توسط رگرسیون، در شکل ۱ آمده است. سپس به کمک رابطه به دست آمده از رگرسیون و مقادیر چاه‌نگارها، برای بررسی قابلیت تعمیم روش، تخلخل در یک چاه دیگر تخمین زده شده است (شکل ۲). همانطور که در شکل ۲ مشاهده می‌شود، همبستگی بین مقادیر تخلخل تخمینی توسط رگرسیون خطی چندگانه با مقادیر تخلخل مغزه در چاه مورد مطالعه در حدود ۰/۸ می‌باشد.



شکل ۱. همبستگی تخلخل حاصل از رگرسیون چندگانه و مغزه برای داده‌های ۴ چاه مورد مطالعه.



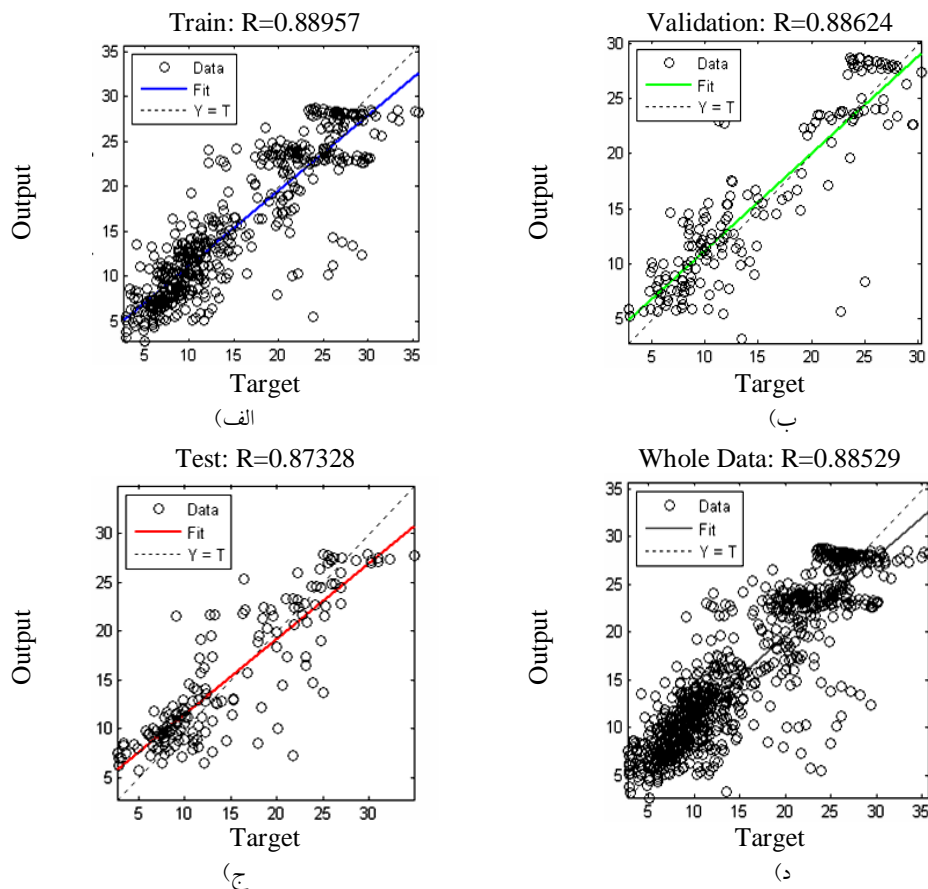
شکل ۲. همبستگی تخلخل حاصل از رگرسیون چندگانه و مغزه در حالت بررسی تعمیم‌پذیری روش در چاه دیگر.

۲-۷ شبکه عصبی چندلایه

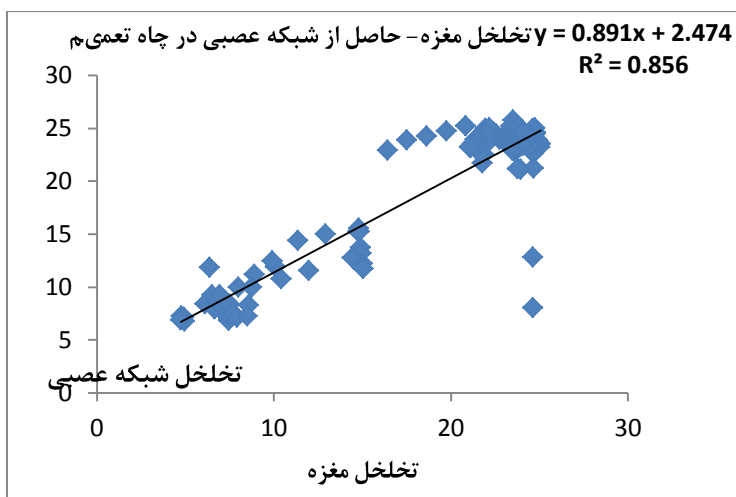
بر خلاف حالت قبل شبکه‌های عصبی، مدل‌های رگرسیون پیچیده‌ی غیرخطی هستند؛ که ساختار آنها به صورت تجربی معین می‌شود. متغیرهای ورودی جهت آموزش شبکه، مقادیر چاه‌نمودارها در ۴ چاه از میدان مورد مطالعه، و خروجی شبکه مقادیر تخلخل مغزه می‌باشد. شبکه بهینه انتخاب شده برای تخمین تخلخل با متغیرهای ورودی ذکر شده و ضریب همبستگی بین

تخلخل واقعی و تخمین زده شده توسط شبکه عصبی در چهار حالت آموزش، اعتبارسنجی، آزمون و کل داده‌ها در شکل ۳ نشان داده شده است. این شبکه دارای ۳ لایه میانی است، که در لایه میانی اول ۱۰ نرون، در لایه میانی دوم ۸ نرون و در لایه میانی سوم ۶ نرون وجود دارد. هر کدام از این نمودارها، دارای خط برازش شده‌ای بین پاسخ شبکه و داده‌های چاه نمودار هستند. این خط دارای یک شیب و ضریب همبستگی است که پراکندگی داده‌ها را حول خط برازش شده نشان می‌دهد.

هر چه مقدار ضریب همبستگی و شیب خط به عدد ۱، نزدیک‌تر و پراکندگی داده‌ها حول خط برازش کمتر باشد، نشان از پیش‌بینی بهتر شبکه در تخمین تخلخل دارد. همانند حالت قبل برای بررسی تعمیم‌پذیری روش، از داده‌های یک چاه دیگر جهت تخمین تخلخل استفاده شده است (شکل ۴). همان‌طور که در شکل ۴ مشاهده می‌شود، همبستگی بین مقادیر تخلخل تخمینی توسط شبکه عصبی با مقادیر تخلخل مغزه در حدود ۰/۸۵ می‌باشد.



شکل ۳. همبستگی بین تخلخل واقعی و تخمین زده شده توسط شبکه عصبی در سه حالت، الف) آموزش ب) اعتبارسنجی ج) آزمون و د) کل داده‌ها برای داده‌های ۴ چاه مورد مطالعه.



شکل ۴. همبستگی بین تخلخل واقعی و تخمین زده شده توسط شبکه عصبی در حالت بررسی تعمیم پذیری روش در چاه دیگر.

۸. تخمین تخلخل با رویکرد ترکیب اطلاعات

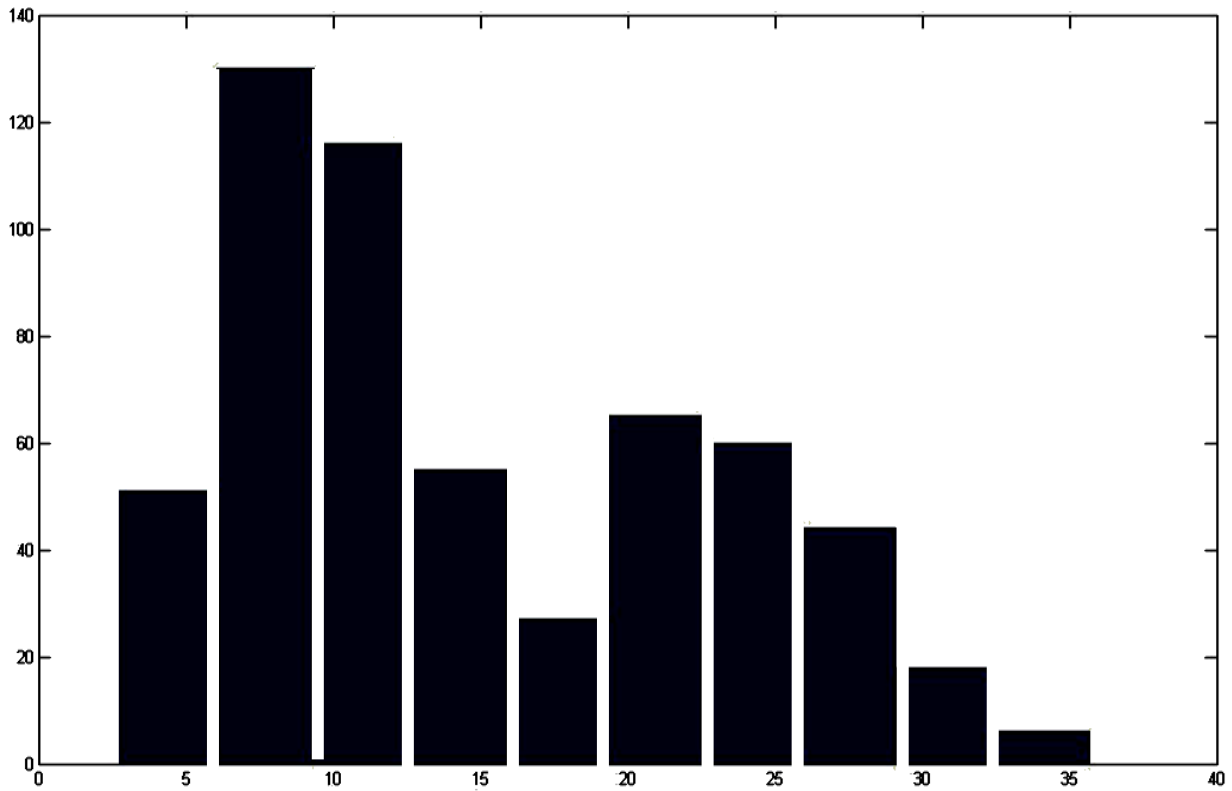
متغیرهای ورودی مورد استفاده در این تحقیق، شامل چاه‌نمودارهای صوتی (DT)، نوترون ($NPHI$)، چگالی ($RHOB$)، مقاومت ویژه لئولوگ عمیق (LLD)، مقاومت ویژه لئولوگ کم عمق (LLS)، پرتو گاما (CGR) و مقاومت ویژه کروی میکرو ($MSFL$) هستند. در نتیجه، حداکثر از هفت متغیر ورودی جهت تخمین تخلخل به روش تئوری بیزین استفاده شده است. در ادامه مراحل که در این تحقیق برای تخمین تخلخل با استفاده از تئوری بیزین به کار رفته، آمده است.

- استفاده از داده‌های چاه‌نمودارها و مغزه چهار چاه از میدان مورد مطالعه.
- تقسیم‌بندی داده‌های این چهار چاه به ۷۰٪ داده آموزش و ۳۰٪ داده آزمون.
- طراحی طبقه‌بندی کننده بیزین بر مبنای تعداد کلاس‌ها و متغیرهای ورودی در حالت‌های مختلف.
- کلاس بندی بر مبنای هیستوگرام توزیع فراوانی ۷۰٪ داده‌های تخلخل مغزه (داده‌های آموزش) که نمودار آن در شکل ۵ آمده است، در حالات ۲، ۳، ۴، ۵ و ۶ کلاس با در نظر گرفتن ۳، ۴، ۵، ۶ و ۷ متغیر ورودی برای هر یک از کلاس‌های ذکر شده.

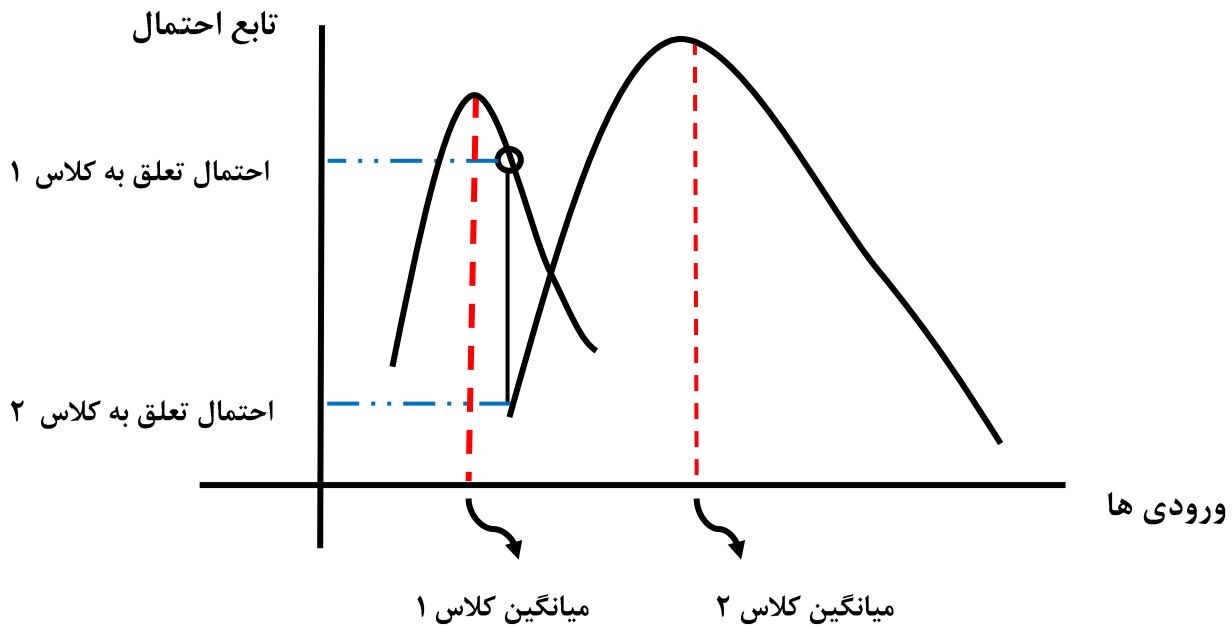
- تخمین تابع توزیع احتمال در هر یک از کلاس‌ها و محاسبه متغیرهای مربوط به آنها، یعنی میانگین و کوواریانس.
 - محاسبه احتمال درجه‌ی تعلق هر داده به هر یک از کلاس‌ها با استفاده از رابطه‌ی بیزین.
 - تخمین تخلخل با استفاده از رابطه‌ی امید ریاضی.
 - تعیین حالات بهینه تخمین از لحاظ تعداد کلاس و متغیرهای ورودی.
- رابطه‌ی امید ریاضی برای محاسبه‌ی تخلخل در حالت ۲ کلاس نشان داده شده در شکل ۶، به صورت زیر است:
- $$\text{امید ریاضی} = \text{مقدار تخلخل تخمینی} = \text{میانگین کلاس ۱} * \text{احتمال تعلق به کلاس ۱} + \text{میانگین کلاس ۲} * \text{احتمال تعلق به کلاس ۲}$$

در نهایت، حالات بهینه تخمین از لحاظ تعداد کلاس و متغیرهای ورودی مشخص شد. به عنوان مثال نمونه‌ای از بهترین نتایج حاصل از تخمین به روش بیزین در چاه‌های مورد مطالعه در جدول ۳ آمده است. با بررسی شکل‌های حاصل از تخمین بیزین، حالات بهینه (جدول ۴) انتخاب شد. نمونه‌هایی از تخمین‌های بهینه در نمودارهای شکل ۷ آمده است. همان‌طور که در این شکل دیده می‌شود، هر کدام از این نمودارها، دارای خط برازش شده‌ای بین پاسخ حاصل از تکنیک ترکیب اطلاعات (تئوری بیزین) و داده‌های چاه‌نمودارها هستند. این خط دارای یک شیب و ضریب همبستگی است که پراکندگی داده‌ها را

حول خط برازش نشان می‌دهد. هر چه مقدار ضریب همبستگی و شیب خط به عدد ۱ نزدیک‌تر و پراکندگی داده‌ها حول خط برازش کمتر باشد، نشان از پیش‌بینی بهتر این تکنیک دارد. با توجه به ضریب همبستگی بالای بین مقادیر تخمین زده شده و واقعی، به خوبی می‌توان دریافت که تکنیک ترکیب اطلاعات (تئوری بیزین) در تخمین تخلخل موفق بوده است.



شکل ۵. هیستوگرام تخلخل بر مبنای ۷۰ درصد داده‌های تخلخل مغزه (آموزش) در چاه مورد مطالعه.



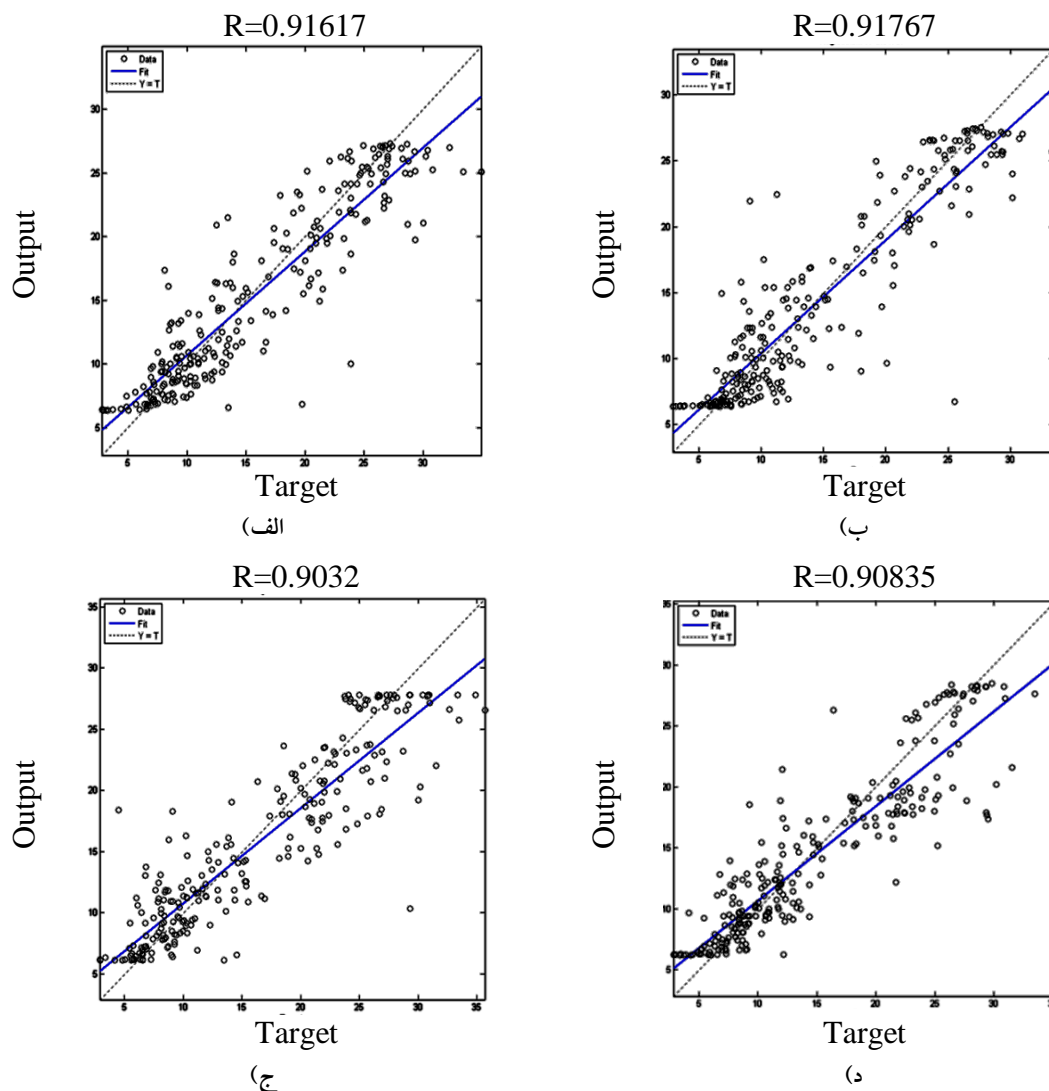
شکل ۶. استفاده از امید ریاضی برای محاسبه تخلخل در حالت ۲ کلاس.

جدول ۳. نمونه‌ای از بهترین نتایج حاصل از تئوری بی‌زین برای تخمین تخلخل در ۴ چاه مورد مطالعه.

ضریب تعیین (R^2)	حالت مورد بررسی	
	کلاس	متغیر
٪۰.۶۷	۲	۳
٪۰.۷۷	۳	۳
٪۰.۷۹	۴	۳
٪۰.۷۹	۵	۳
٪۰.۸۳	۶	۳
٪۰.۷۶	۲	۴
٪۰.۷۷	۳	۴
٪۰.۷۷	۴	۴
٪۰.۷۹	۵	۴
٪۰.۸۳	۶	۴
٪۰.۷۶	۲	۵
٪۰.۷۷	۳	۵
٪۰.۷۷	۴	۵
٪۰.۸۱	۵	۵
٪۰.۷۹	۶	۵
٪۰.۷۲	۲	۶
٪۰.۷۷	۳	۶
٪۰.۸۳	۴	۶
٪۰.۸۱	۵	۶
٪۰.۸۱	۶	۶
٪۰.۷۷	۲	۷
٪۰.۸۳	۳	۷
٪۰.۸۱	۴	۷
٪۰.۸۶	۵	۷
٪۰.۸۳	۶	۷

جدول ۴. حالات بهینه تخمین از لحاظ تعداد کلاس و متغیرهای ورودی در ۴ چاه مورد مطالعه.

ضریب تعیین (R ²)	حالت مورد بررسی	
	کلاس	متغیر
٪۷۹	۵	۳
٪۸۳	۶	۳
٪۷۷	۴	۴
٪۷۹	۵	۴
٪۸۳	۶	۴
٪۷۹	۶	۵
٪۸۳	۴	۶
٪۸۱	۵	۶
٪۸۱	۶	۶
٪۸۱	۴	۷
٪۸۶	۵	۷
٪۸۳	۶	۷



شکل ۷. همبستگی بین تخلخل واقعی و تخمین زده شده توسط تئوری بیزین در سه حالت. الف) ۳ متغیر و ۶ کلاس ب) ۴ متغیر و ۶ کلاس ج) ۶ متغیر و ۵ کلاس و د) ۶ متغیر و ۶ کلاس برای داده‌های آزمون ۴ چاه مورد مطالعه.

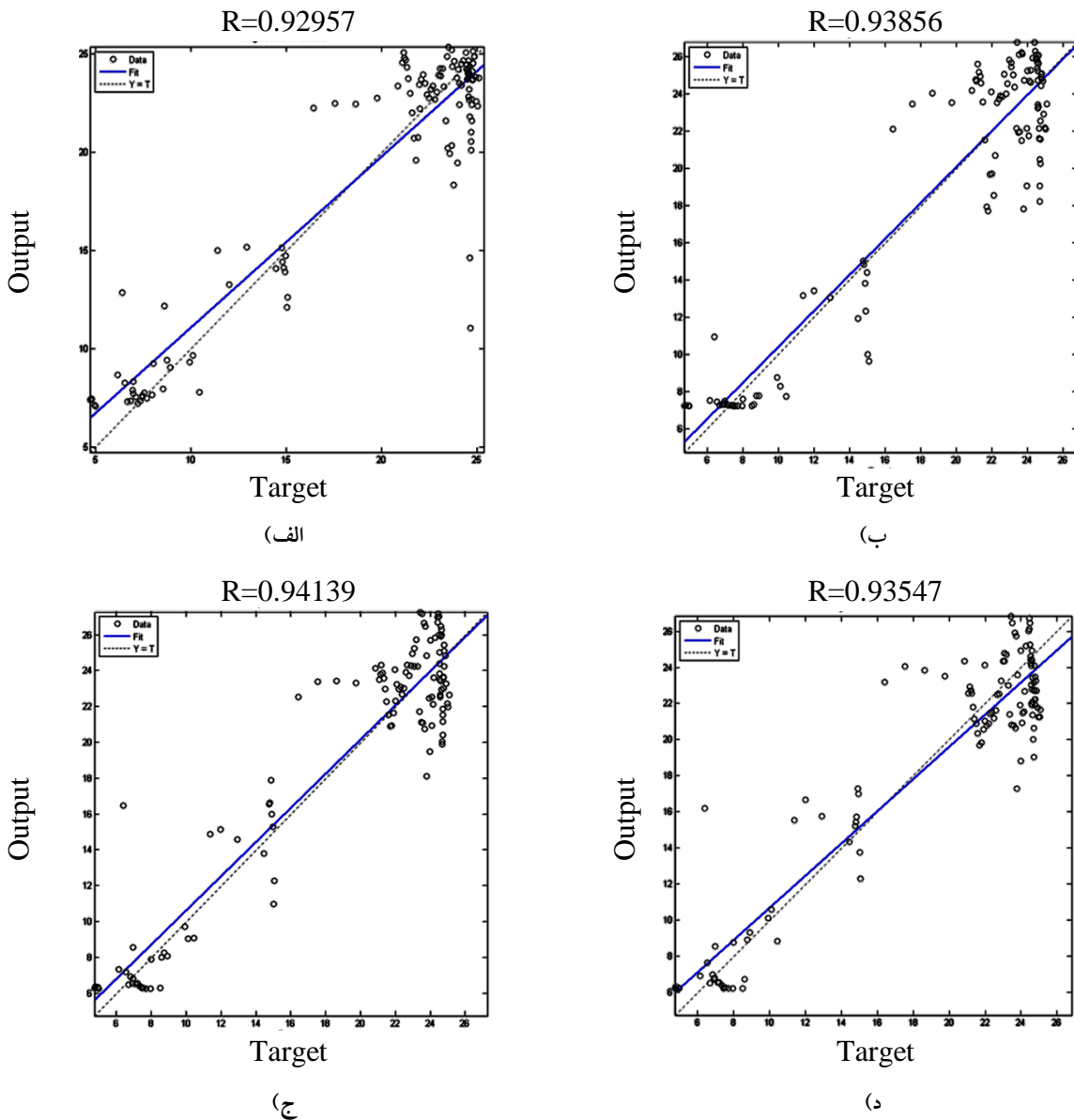
حال برای بررسی قابلیت تعمیم تخمین روش بیزین، از داده‌های یک چاه دیگر برای تخمین تخلخل استفاده شد؛ که نتایج این بررسی در جدول ۵ آمده است. با بررسی شکل‌های حاصل، حالات بهینه تخمین انتخاب شد، که نتایج بهینه در جدول ۶ آمده است. نمونه‌ای از تخمین‌های بهینه نیز در شکل ۸ نشان داده شده است. در این شکل همانند حالت آموزش، با توجه به ضریب همبستگی بالا (بیش از ۰/۹) حاصل بین مقادیر تخمین زده شده و واقعی، در حالت بررسی تعمیم‌پذیری روش، به خوبی می‌توان دریافت که تکنیک ترکیب اطلاعات (تئوری بیزین) در تخمین تخلخل موفق ظاهر شده است. در انتها تخمین حاصل از ۳ روش رگرسیون خطی چندگانه، شبکه عصبی چندلایه و ترکیب اطلاعات (تئوری بیزین) در حالت تعمیم‌پذیری در چاه مورد مطالعه، با یکدیگر مقایسه شد. نتیجه آنکه تکنیک ترکیب اطلاعات (تئوری بیزین) نسبت به دو روش رگرسیون خطی چندگانه و شبکه عصبی در تخمین تخلخل به مقدار قابل توجهی بهتر عمل نموده است (جدول ۷).

جدول ۵. نمونه‌ای از بهترین نتایج حاصل در بررسی تعمیم‌پذیری تخمین به روش بیزین در ۴ چاه مورد مطالعه.

ضریب تعیین (R^2)	حالت مورد بررسی	
	کلاس	متغیر
٪۸۳	۲	۳
٪۸۳	۳	۳
٪۸۵	۴	۳
٪۸۳	۵	۳
٪۸۵	۶	۳
٪۸۵	۲	۴
٪۸۳	۳	۴
٪۸۳	۴	۴
٪۸۵	۵	۴
٪۸۳	۶	۴
٪۸۶	۲	۵
٪۸۶	۳	۵
٪۸۵	۴	۵
٪۸۳	۵	۵
٪۸۱	۶	۵
٪۸۶	۲	۶
٪۸۸	۳	۶
٪۸۶	۴	۶
٪۸۸	۵	۶
٪۸۱	۶	۶
٪۹۰	۲	۷
٪۸۵	۳	۷
٪۸۶	۴	۷
٪۸۶	۵	۷
٪۷۷	۶	۷

جدول ۶. حالات بهینه تخمین از لحاظ تعداد کلاس و متغیرهای ورودی در بررسی تعمیم پذیری تخمین به روش بیزین در چاه مورد مطالعه.

ضریب تعیین (R^2)	حالت مورد بررسی	
	کلاس	متغیر
٪۸۳	۳	۳
٪۸۵	۴	۳
٪۸۳	۵	۳
٪۸۳	۳	۴
٪۸۵	۵	۴
٪۸۶	۳	۵
٪۸۵	۴	۵
٪۸۸	۳	۶
٪۸۶	۴	۶
٪۸۸	۵	۶
٪۸۶	۴	۷
٪۸۶	۵	۷



شکل ۸. همبستگی بین تخلخل واقعی و تخمین زده شده توسط تئوری بیزین در سه حالت، الف) ۳ متغیر و ۴ کلاس ب) ۶ متغیر و ۴ کلاس ج) ۶ متغیر و ۵ کلاس و د) ۷ متغیر و ۵ کلاس برای چاه تعمیم.

جدول ۷. مقایسه ضرایب رگرسیون تخمین با ۳ تکنیک رگرسیون خطی چندگانه، شبکه عصبی برگشت پذیر خطا و ترکیب اطلاعات (تئوری بیزین) در حالات تعمیم پذیری ۳ روش در چاه مورد مطالعه.

تکنیک مورد بررسی جهت تخمین تخلخل	ضریب تعیین (R^2)
رگرسیون خطی چندگانه	۰/۸۳
شبکه عصبی چندلایه	۰/۸۵
ترکیب اطلاعات (تئوری بیزین، با ۶ متغیر ورودی و ۵ کلاس)	۰/۸۸

۹. نتیجه گیری

در این تحقیق، برای تعیین مقادیر تخلخل در چاه مورد مطالعه واقع در میدان نفتی آزادگان، از دو تکنیک شبکه عصبی چندلایه و رگرسیون خطی چندگانه استفاده شد و در نهایت، نتایج این دو روش با تکنیک ترکیب اطلاعات (تئوری بیزین) مقایسه شد. با بررسی مقادیر تخلخل تخمین زده شده در حالات تعمیم پذیری هر یک از ۳ روش مذکور، افزایش مقدار ضریب تعیین تا حد ۸۸٪ در حالت ۶ متغیر ورودی و ۵ کلاس تکنیک ترکیب اطلاعات (تئوری بیزین)، مشاهده شد. یعنی ضریب تعیین با تئوری بیزین نسبت به تخمین با شبکه عصبی ۳٪ و نسبت به رگرسیون خطی چندگانه ۵٪ افزایش را نشان داد. هم چنین با استفاده ترکیب اطلاعات بیزین می توان، با تعداد متغیرهای ورودی (چاه نمودارهای) کمتر، به ضرایب همبستگی بالاتری نسبت به دو روش شبکه عصبی و رگرسیون خطی چندگانه دست یافت. از این رو تکنیک ترکیب اطلاعات (تئوری بیزین) در تخمین تخلخل به مقدار قابل توجهی از دو تکنیک شبکه عصبی چندلایه و رگرسیون خطی چندگانه بهتر عمل نمود و در این حالت ضرایب رگرسیون بالایی بین مقادیر تخلخل تخمین زده شده و واقعی حاصل شد. در نتیجه تکنیک های ترکیب اطلاعات از جمله تئوری بیزین می توانند به عنوان ابزاری ساده، ارزان و قدرتمند به منظور تخمین تخلخل به کار روند.

"هیئت تحریریه مجله از آقایان دکتر علیرضا عرب امیری و دکتر محمد کنشلو که داوری مقاله را بر عهده داشته اند

کمال تشکر و سپاس را دارد"

منابع

- [1]- Tarek, A., 1946. "Reservoir rock properties and fluid flow in porous media" London.
- [2]. M. Jalali Lichaei & Dr. M. Nabi – Bidhendi., 2006- "Comparison between Multiple Linear Regression and Artificial Neural Networks for Porosity and Permeability Estimation"
- [3]. Rezaee M.R., Kadkhodaie A., Barabadi A., (2007), "Prediction of shear wave velocity from petrophysical data utilizing intelligent systems: An example from a sandstone reservoir of Carnarvon Basin, Australia", Journal of Petroleum Science and Engineering, 55, pp. 201–212.
- [4]. Luo, R. C. and Kay, M. G. (1995). "Multisensor integration and fusion for intelligent machines and system", Ablex Publishing Corp.
- [5]. Abidi, M. A. and Gonzalez, R. C. (1992). "Data fusion in robotics and machine intelligence."
- [6]. Hall, D.L. and Llinas, J., 2001. "Handbook of multisensor data fusion." CRC Press LLC, Boca Raton
- [7]. Worboys, M. and Duckham, M. (2004). GIS: "A Computing Perspective", CRC Press, Lon
- [8]. Djebbar Tiab and Erle C. Donaldson, "Petrophysics"
- [9]. Balan, B., Mohaghegh, S., Ameri, S., 1995- "State - of -Art in permeability determination from well log data : Part - 1 -A , comparative study , Model development." SPE 30978 , PP:17-25.
- [10]. Balan, B., Mohaghegh, S., Ameri, S., 1995- "State - of -Art in permeability determination from well log data : Part - 1 -A , comparative study , Model development." SPE 30978 , PP:17-25.
- [11]. Mohaghegh S.D., (2000), "Virtual Intelligence and its Applications in Petroleum Engineering, 1. Artificial Neural Networks", Journal of Petroleum Technology, Distinguished Author Series.
- [12]. Helle H.B., Bhatt A.I and Ursin B., (2001), "Porosity and permeability prediction from wireline logs using artificial neural networks: a North Sea case study", Geophysical Prospecting, 49, pp. 431- 444
- [13]. Dasarathy, B.V., 1991. "Decision fusion strategies in multisensor environments Systems", Man and Cybernetics, 21(5): 1140-1154.
- [14]. Hall, D. L. and J. Llinas (2001). Handbook of multisensor data fusion. Boca Raton, CRC Press LLC.
- [15]. Tahani, H. and J. M. Keller (1990). "Information fusion in computer vision using the fuzzy integral. "

[16]. Russo, F. and G. Ramponi (1994). "Fuzzy methods for multisensor data fusion." *Instrumentation and Measurement, IEEE Transactions on* 43(2): 288-294.

[17]. Gros, X.E., 1997, "NDT Data Fusion", John Wiley & Sons, New York, 205 pages.

[18]. B. Huang, "Bayesian methods for control loop monitoring and diagnosis," *Jurnal of process Control*, Vol. 18, pp. 829-838, Oct 2008.

[19]. James lines, Dheeraj Talreja, Christopher Bowman, June 2004. Design of a Performance Evaluation Methodology for Data Fusion – based Multiple Target Tracking Systems": Part 2: Analysis Methodologies.

[۲۰]. عاقبتی، رضا، "معرفی یک میدان- طرح توسعه میدان مورد مطالعه"، نشریه فنی تخصصی شرکت ملی نفت ایران- اکتشاف و تولید، شماره ۵۱، آبان ماه ۱۳۸۷

Porosity modeling in Azadegan oil field: a comparative study of Bayesian theory of data fusion, multi layer neural network, and multiple linear regression techniques

Mazaheri T. A., Memarian H., Thokhmchi B., Moshiri B.

Abstract

Porosity parameter is an important reservoir property that can be obtained by studying the well core. However, all wells in a field do not have a core. Additionally, in some wells such as horizontal wells, measuring the well core is practically impossible. However, for almost all wells, log data is available. Usually these logs are used to estimate porosity. The porosity value obtained from this method is influenced by factors such as temperature, pressure, fluid type, and amount of hydrocarbons in shale formations. Thus it is slightly different from the exact value of porosity. Thus, estimates are prone to error and uncertainty. One of the best and yet most practical ways to reduce the amount of uncertainty in measurement is using various sources and data fusion techniques. The main benefit of these techniques is that they increase confidence and reduce risk and error in decision making. In this paper, in order to determine porosity values, data from four wells located in Azadegan oil field are used. First, multilayer neural network and multiple linear regressions are used to estimate the values and then the results of these techniques are compared with a data fusion method (Bayesian theory). To check if it would be possible to generalize these three methods on other data, the porosity parameter of another independent well in this field is also estimated by using these techniques. Number of input variables to estimate porosity in both the neural network and the multiple linear regressions methods is 7, and in the data fusion technique, a maximum of 7 input variables is used. Finally, by comparing the results of the three methods, it is concluded that the data fusion technique (Bayesian theory) is a considerably more accurate technique than multilayer neural network, and multiple linear regression, when it comes to porosity value estimation; Such that the results are correlated with the ground truth greater than 90%.

Keywords: uncertainty, estimation, data fusion, Bayesian theory, multi layer neural network, multiple linear regression, Azadegan oil field, Sarvak formation, Iran.