



## پیش بینی درصد اشباع آب سازند با استفاده از داده‌های پتروفیزیکی با تمرکز روی گوناگونی خبره‌ها

سید علی جعفری کناری\*، مرکز آموزش صنعت نفت محمودآباد، شرکت ملی نفت ایران

چکیده

پیش بینی پارامترهای پتروفیزیکی مخازن با استفاده از تکنیک یادگیری دسته‌جمعی، توجه بسیاری از محققین را به خود جلب نموده است. در تمامی فعالیت‌های انجام شده با روش مذکور، انتخاب خبره‌های مجموعه یادگیری، فقط بر اساس عملکرد و دقت نتایج خروجی آن‌ها انجام گردیده است. لذا، یکی از نکاتی که در استفاده از این روش تاکنون به آن توجه نشده، گوناگونی خبره‌های تشکیل دهنده مجموعه یادگیری فوق است که تا حد زیادی در بهبود و موفقیت نتایج پیش‌بینی مؤثر است. از این رو، هدف این مقاله اهتمام به این پارامتر و تمرکز روی آن در انتخاب اعضای مجموعه یادگیری دسته‌جمعی است. در این راستا، مجموعه‌ای متشکل از تعداد ۱۰۰ خبره از نوع شبکه عصبی مصنوعی با روش یادگیری LM و با دستکاری در داده‌های ورودی، تشکیل شده است. سپس چهار روش اندازه‌گیری گوناگونی خبره‌ها معرفی و با استفاده از آنها، خبره‌های با گوناگونی بالاتر جهت عضویت در مجموعه هرس شده‌اند. نتایج حاصل از اعمال روش پیشنهادی، روی داده‌های چاه‌پیمایی جهت پیش‌بینی اشباع آب مخزن تحت مطالعه، نشان‌دهنده موفقیت عملکرد آن در مقایسه با مجموعه یادگیری اولیه است.

واژگان کلیدی یادگیری دسته‌جمعی، گوناگونی بین خبره‌ها، شبکه‌های عصبی مصنوعی، مخازن نفت، پارامترهای پتروفیزیکی، درصد اشباع آب

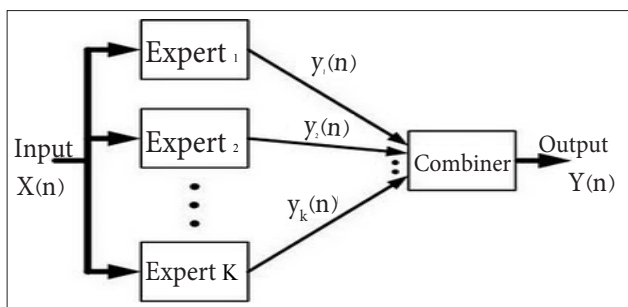
### مقدمه

شبکه‌های ضعیف‌تر از دست بروند. به عبارت دیگر، مدل‌های دور ریز شده، ممکن است برای برخی از نمونه‌ها پیش‌بینی بهتری داشته باشند، حال آنکه حذف شده‌اند [۲، ۱]. لذا، این روش به دلیل از دست دادن الگوی نهفته شده در داده‌های تحت آزمایش برخی ناحیه‌ها، نمی‌تواند در یادگیری کامل روابط بین داده‌ها مؤثر باشد. ایرادهای ذکر شده باعث ایجاد انگیزه‌ای برای معرفی تکنیک یادگیری دسته‌جمعی<sup>۱</sup> یا کمیته‌ای از ماشین‌های یادگیری شده است. ترکیب ماشین‌های<sup>۱۰</sup> یادگیری یا مفهوم یادگیری دسته‌جمعی، برای اولین بار در کار نیلسون (۱۳۶۵) ظاهر شد [۳] و پس از آن، مطالعات گسترده‌ای در این رابطه از دهه ۱۹۹۰ آغاز گردید. امروزه کارهای زیادی در خصوص استفاده از این متد، جهت تعیین پارامترهای مخازن نفت و گاز انجام گرفته است [۱، ۲، ۴-۱۰]. یکی از محدودیت‌های اصلی در مطالعات مذکور، آن است که انتخاب خبره‌ها تنها بر اساس عملکرد و دقت خروجی آنهاست. بدین ترتیب که در ابتدا، نوع خبره‌ها انتخاب و سپس مدل‌ها یک‌به‌یک تحت آموزش قرار می‌گیرند. برای یک مدل مشخص، آنقدر عمل یادگیری تکرار می‌شود تا بالاترین عملکرد و دقت به دست آید، آنگاه مدل فوق به مجموعه یادگیری دسته‌جمعی اضافه می‌شود. در نتیجه، نکته قابل اهمیت در این

استفاده از روش‌های هوشمند برای پیش‌بینی پارامترهای پتروفیزیکی مخازن، طی دو دهه اخیر مورد توجه مهندسان نفت قرار گرفته است. از مهمترین آنها می‌توان به متدهایی نظیر شبکه‌های عصبی مصنوعی<sup>۱</sup>، تئوری فازی<sup>۲</sup>، شبکه نروفازی<sup>۳</sup> و الگوریتم ژنتیک<sup>۴</sup> اشاره کرد. در کنار محدودیت‌هایی که هر یک از این تکنیک‌های یادگیری ماشین داشته و دارند، به کارگیری آن‌ها کمک بسیار زیادی به شناسایی و برآورد مشخصات مخازن نفت کرده است. یکی از ایرادهایی که به‌نحوه استفاده و به کارگیری این ابزارها برمی‌گردد، آن است که یک سیستم هوشمند به‌تنهایی نمی‌تواند تعمیم قابل اطمینانی برای داده‌های تحت آزمایش را ارائه کند. به عبارت دیگر، از آنجایی که انتخاب مدل نهایی پس از چندین بار تکرار مرحله یادگیری و بر اساس میزان دقت<sup>۵</sup> و عملکرد<sup>۶</sup> آن به‌زای داده‌های آموزش<sup>۷</sup> انجام می‌پذیرد، لذا در هنگام به کارگیری آن برای داده‌های تست<sup>۸</sup>، هیچ تضمینی برای دریافت یک برآورد مناسب از داده هدف وجود ندارد. در واقع، انتخاب بهترین پیش‌بینی کننده از میان شبکه‌های آموزش دیده، لزوماً انتخاب ایده‌آل نیست؛ به این معنی که اطلاعاتی که به‌طور بالقوه با ارزش هستند، ممکن است با حذف

\* نویسنده عهده‌دار مکاتبات (s.a.jafari@nioc.ir)

بوده است. لذا، در این مقاله با هدف تعیین پارامترهای مخزن با استفاده از تکنیک یادگیری دسته جمعی، تلاش می‌گردد روش‌هایی که تاکنون بکار گرفته شده‌اند، بهبود یابد. طراحی اولیه تکنیک یادگیری که آموزش خبره‌ها به صورت مستقل از هم باشند، در ابتدا شامل دو مرحله طراحی و ایجاد خبره و سپس اعمال متد ترکیبی مناسب روی خروجی خبره‌ها بوده است (شکل-۱). یکی از معایب چنین رویکردی، از دست دادن تعامل بین اعضای مجموعه در طول فرآیند یادگیری است که هیچ بازخوردی از مرحله ترکیب به مرحله طراحی اعضا وجود ندارد. بنابراین، ممکن است که برخی از خبره‌ها، هیچ یا سهم اندکی را در افزایش دقت کل مجموعه داشته باشند. محققین بسیاری، افزودن یک مرحله میانی به این متد را پیشنهاد کرده‌اند که آن را مرحله هرس کردن<sup>۲۲</sup> نام نهاده‌اند [۲۱-۲۷]. در این مرحله، اعضای مجموعه اولیه، غربال شده و در نهایت، به تجربه، حداکثر ۳۰-۲۰ درصد از اعضای اصلی، جهت عضویت در مجموعه نهایی باقی می‌مانند. هرس کردن مجموعه اولیه متشکل از خبره‌های آموزش دیده، در حالی که گوناگونی و تنوع بالایی را در میان اعضای باقیمانده مجموعه فوق عرضه می‌کند، یک روش موثر برای افزایش دقت پیش‌بینی مورد نظر است [۲۸، ۲۹]. بنابراین هرس کردن مجموعه فوق، شبیه به یک مسأله بهینه‌سازی است که در آن، هدف، یافتن زیرمجموعه‌ای بهینه از خبره‌های مجموعه اصلی است، به گونه‌ای که باعث بهبود در کارایی مجموعه نهایی گردد. بدین سبب، در این مقاله، ابتدا تعدادی خبره با استفاده از شبکه عصبی مصنوعی تولید شده، سپس، با استفاده از متدهای مختلف اندازه‌گیری گوناگونی خبره‌ها، به هرس کردن مجموعه اولیه و در نهایت انتخاب اعضای نهایی پرداخته می‌شود. در انتها نیز، عملکرد روش پیشنهادی بر روی داده‌های استفاده شده در مقاله [۱۸] جهت پیش‌بینی درصد اشباع آب سازند با استفاده از داده‌های پتروفیزیکی بررسی می‌گردد. نتایج حاصل نشان می‌دهد که مجموعه نهایی هرس شده دارای دقت پیش‌بینی بهتری نسبت به مجموعه اولیه یادگیری ماشین بوده و علاوه بر آن، از نظر میزان حافظه مورد نیاز و



۱ | یک نمونه از یادگیری دسته جمعی دو مرحله‌ای با K خبره

زمینه که به آن توجه نشده است، مفهوم گوناگونی<sup>۱۱</sup> در بین اعضای یک مجموعه یادگیری دسته جمعی است که به متمایز بودن خطاهای بین خبره‌ها تعبیر می‌شود [۱۲، ۱۱]. نتایج تحقیقات تئوری [۱۳، ۱۴] و تجربی [۱۵] نشان می‌دهد که موفقیت روش یادگیری دسته جمعی تا حد زیادی بستگی به تنوع اعضای دربرگیرنده آن دارد. پژوهش‌هایی که تاکنون جهت پیش‌بینی پارامترهای مختلف مخازن (صرف نظر از نوع پارامتر) با استفاده از یادگیری دسته جمعی انجام پذیرفته است را می‌توان به دو دسته عمده همگن<sup>۱۲</sup> و ناهمگن<sup>۱۳</sup> تقسیم کرد. در مدل‌های همگن، خبره‌ها با تغییرات در پارامترهای مختلف یک مدل یادگیری هوشمند مفروض ایجاد می‌گردند، نظیر شرایط اولیه یا توقف، تکنیک‌های نمونه‌برداری<sup>۱۴</sup> یا نمونه‌برداری مجدد<sup>۱۵</sup>، دستکاری<sup>۱۶</sup> در نمونه‌های یادگیری، دخل و تصرف در ویژگی‌های ورودی و خروجی. در مقابل، یک گروه ناهمگن با خبره‌هایی که از مدل‌های یادگیری مختلف تولید شده‌اند، ساخته می‌شود. تحقیقات انجام شده [۱، ۲، ۴، ۵، ۷، ۱۶] نمونه‌هایی از به کارگیری متد همگن و همچنین برخی دیگر [۴، ۱۷، ۱۸] نمونه‌هایی از به کارگیری متد ناهمگن در پیش‌بینی پارامترهای مختلف مخازن نفت را نشان می‌دهند. در همه روش‌های فوق، هر یک از خبره‌های کاندید ابتدا تحت آموزش متوالی قرار گرفته و سپس، در تکراری که دارای بالاترین دقت است، متوقف شده و نهایتاً، خبره آموزش دیده فوق به مجموعه یادگیری دسته جمعی اضافه می‌گردد. همچنین، تعداد اعضای این مجموعه به صورت دلخواه و توسط کاربر و تنها با توجه به در دسترس بودن متدها یا اهمیت آنها تعیین می‌شوند. لازم به ذکر است که اگر تمام خبره‌های تولید شده دارای دقت بالا باشند، آنگاه جمع‌آوری آنها برای تشکیل مجموعه فوق، تأثیری در بهبود دقت نتیجه نهایی نداشته و بنابراین، نیازی به بررسی گوناگونی بین اعضا نیست. لذا، تمرکز مقاله حاضر در تولید مجموعه یادگیری دسته جمعی با اهتمام به گوناگونی بین اعضا و تاکید بر اهمیت آن در بهبود عملکرد و دقت پیش‌بینی نهایی مقادیر هدف است.

سه روش عمده برای طراحی یک مجموعه یادگیری دسته جمعی با اهتمام به گوناگونی اعضای مجموعه وجود دارد که خبره‌ها به صورت مستقل<sup>۱۷</sup> از هم، پی‌درپی<sup>۱۸</sup> و یا به‌طور همزمان<sup>۱۹</sup> تحت آموزش قرار می‌گیرند که البته دو روش آخر، مورد بحث این مقاله نیست [۱۲، ۱۵، ۱۹، ۲۰]. دلیل اتخاذ چنین تصمیمی آن است که تمام کارهایی که تاکنون در صنعت نفت در جهت تعیین پارامترهای مخزن با روش مورد بحث (یادگیری دسته جمعی) انجام گرفته، به دنبال ایجاد خبره‌های گوناگون نبوده و تنها به دنبال تولید و ترکیب نهایی آنها با استفاده از روش‌های میانگین‌گیری ساده<sup>۲۰</sup> و میانگین‌گیری موزون<sup>۲۱</sup>



سرعت محاسبات نیز، عملکرد بهتری را ارائه می‌دهد.

و  $Y^n = \{y_1^n, y_2^n, \dots, y_N^n\}$

### ۱- تاریخچه

تکنیک‌های هوشمند مختلفی برای پیش‌بینی اشباع آب سازند با استفاده از داده‌های پتروفیزیکی به کار گرفته شده است. البلوشی و همکاران (۲۰۰۹) با استفاده از شبکه‌های عصبی مصنوعی و به کارگیری داده‌های چاه‌پیمایی و داده‌های مغزه‌ی دین استارک، به پیش‌بینی اشباع آب سازند پرداختند. همچنین، در ادامه، از شبکه عصبی فوق برای تعیین ضریب سیمان و ضریب توان اشباع در معادله آرچی استفاده کردند [۳۰]. آدنیان و همکاران (۲۰۰۹)، روش جدیدی براساس شبکه عملکردی<sup>۲۴</sup> برای تخمین تخلخل و آب اشباع سازند معرفی کردند. تابع استفاده شده جهت پیش‌بینی اشباع آب، تابع لگاریتم بوده که به روش آزمون و خطا برگزیده شده بود. در نهایت، به منظور بررسی اعتبار نتایج حاصل، شبکه عصبی مصنوعی با الگوریتم یادگیری LM مورد استفاده قرار گرفت [۳۱]. علیمردادی و همکاران (۲۰۱۱)، چهار ساختار مختلف با شبکه عصبی مصنوعی جهت پیش‌بینی اشباع آب طراحی کردند که تفاوتشان در توابع انتقال لایه‌های پنهان و خروجی این شبکه بوده است [۳۲]. کمالی و همکاران (۲۰۱۲)، با استفاده از روش شبکه عصبی مصنوعی به پیش‌بینی اشباع آب در یکی از مخازن نفت در جنوب ایران پرداختند. سپس، اشباع آب پیش‌بینی شده توسط مدل پیشنهادی را با مدل‌های مختلف آزمایشگاهی مورد مقایسه و ارزیابی قرار دادند [۳۳]. هله و بات (۲۰۰۲)، با به کارگیری کمیته‌ای از شبکه‌های عصبی مصنوعی به پیش‌بینی درصد اشباع سیالات آب، نفت و گاز پرداختند. از تعداد بیست شبکه‌ی آموزش داده شده، تعداد نه شبکه با بالاترین دقت پیش‌بینی جهت مجموعه یادگیری نهایی برگزیده شدند [۱]. با توجه به ایرادهای بیان شده در بخش مقدمه در خصوص استفاده از یک متد یادگیری ماشین به تنهایی و همچنین کمیته‌ای از ماشین‌های یادگیری بدون توجه به گوناگونی آن‌ها، این مقاله سعی در ارائه روشی دارد که بر هر دو عیب مذکور فائق آید.

### ۲- روش‌های اندازه‌گیری گوناگونی بین خبره‌های یک مجموعه

هیمونتسی (۲۰۰۹)، برخی از متر (شاخص)‌های محاسبه گوناگونی خبره‌های یک مسأله رگرسیون را معرفی کرد [۳۴] که در اینجا به آن اشاره می‌گردد:

فرض کنید داده‌های مورد آزمایش به صورت  $D = \{(x_i, y_i); x_i \in R^n, y_i \in R\}$  و  $Y^m = \{y_1^m, y_2^m, \dots, y_N^m\}$  و  $Y^n = \{y_1^n, y_2^n, \dots, y_N^n\}$  دو بردار شامل مقادیر خروجی وابسته به خبره‌های  $m$  و  $n$  باشد که

### ضریب همبستگی<sup>۲۴</sup>:

$$r = \frac{\sum_{i=1}^N (y_i^m - \mu_{Y^m})(y_i^n - \mu_{Y^n})}{\sqrt{\sum_{i=1}^N (y_i^m - \mu_{Y^m})^2 \sum_{i=1}^N (y_i^n - \mu_{Y^n})^2}} \quad (1)$$

که در آن  $\mu$  معرف میانگین است. در این متر اندازه‌گیری، گوناگونی دو خبره نسبت معکوس با ضریب همبستگی بین آن‌ها دارد. بنابراین، دو خبره با ضریب همبستگی پایین، بر دو خبره با ضریب همبستگی بالا ترجیح داده می‌شود.

### کوواریانس:

$$\text{Cov}(Y^m, Y^n) = E[(Y^m - \mu_{Y^m})(Y^n - \mu_{Y^n})] \quad (2)$$

که در آن  $E$ ، امید ریاضی است. در این متر نیز گوناگونی بین دو خبره با کوواریانس، ارتباط معکوس دارد.

### کای دو<sup>۲۵</sup>:

کای دو  $\chi^2$  نسبت به  $Y^n$  به صورت زیر تعریف می‌شود.

$$\chi^2 = \sum_{i=1}^N \frac{(y_i^m - y_i^n)^2}{y_i^n} \quad (3)$$

در این متر، گوناگونی بین دو خبره با کوواریانس ارتباط مستقیم دارد.

### ناسازگاری<sup>۲۶</sup>:

هو (۱۹۹۸) این متد را اولین بار برای مسائل طبقه‌بندی‌ای ارائه نمود که در آن، هدف، محاسبه میزان ناسازگاری زوجی طبقه‌بندی‌کننده‌هاست [۳۵]. میزان ناسازگاری دو خبره در این روش طبق معادله ۴- اندازه‌گیری می‌شود که در آن، هدف، محاسبه تعداد مشاهدات متمایز دو خبره به تعداد کل مشاهدات است:

$$DIS_{i,j} = \frac{N^{01} + N^{10}}{N^{01} + N^{00} + N^{10} + N^{11}} \quad (4)$$

در این فرمول،  $N$  تعداد نمونه‌ها و یک (۱) و صفر (۰) به ترتیب نشان‌دهنده

۱ | طریقه محاسبه اجزاء فرمول (۴).

	$Y_j \text{ correct}$	$Y_j \text{ wrong}$
$Y_i \text{ correct}$	$N^{11}$	$N^{10}$
$Y_i \text{ wrong}$	$N^{01}$	$N^{00}$

را روی کل مجموعه داده‌ها محاسبه و میزان ناسازگای دو به دوی آن‌ها را از معادله ۴ محاسبه نمود. در این متر نیز، گوناگونی بین دو خبره با کوواریانس ارتباط مستقیم دارد.

### ۳- انتخاب خبره‌های گوناگون یک مجموعه

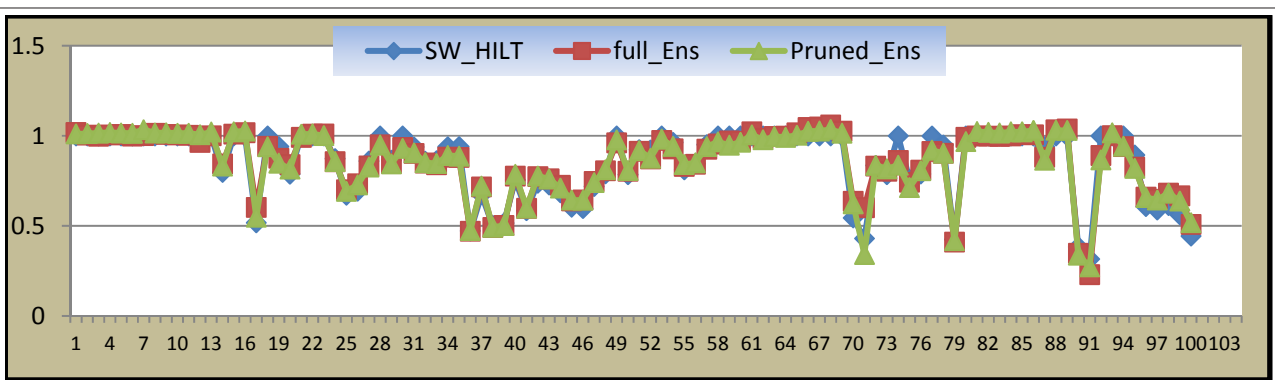
برای انتخاب مجموعه‌ای متنوع و گوناگون از خبره‌ها، دو روش عمده وجود دارد که به روش‌های دوجه‌دو و میانگین معروفند. در روش اول، گوناگونی دوجه‌دوی تمام خبره‌ها اندازه‌گیری شده، سپس، از میان آنها زوج خبره‌هایی که دارای بالاترین مقدار گوناگونی هستند، انتخاب می‌گردد. در روش دوم نیز، میانگین گوناگونی یک خبره نسبت به بقیه خبره‌ها به‌عنوان مقدار گوناگونی آن خبره در نظر گرفته می‌شود.

طبقه‌بندی درست و نادرست است. به عنوان مثال،  $N^0$  معرف تعداد طبقه‌بندی‌هایی است که به وسیله I امین خبره، نادرست و به وسیله J امین خبره، درست طبقه‌بندی شده‌اند (جدول ۱).

برای استفاده از آن در مسائل رگرسیون به صورت زیر عمل می‌کنیم: در ابتدا برای هر داده نمونه  $X$ ، انحراف معیار پیش‌بینی تمام خبره‌ها (یعنی  $\sigma_y$ ) نسبت به ورودی فوق محاسبه می‌شود. حال اگر خروجی اصلی را با  $T$  نمایش دهیم، آنگاه مقدار پیش‌بینی شده توسط خبره I ام اگر در بازه  $[T - \sigma_y, T + \sigma_y]$  قرار گرفت، پس دارای ارزش درست یعنی (۱) و در غیراین صورت، دارای ارزش نادرست یعنی (۰) خواهد بود. حال می‌توان تعداد پیش‌بینی‌های درست و نادرست برای هر خبره

۲ مقایسه عملکرد مجموعه یادگیری اصلی با مجموعه‌های هرس شده از طریق چهار متد تعیین گوناگونی خبره‌ها.

تعداد اعضا مجموعه اولیه	زیر مجموعه هرس شده	همسنگی		کوواریانس		کای دو		ناسازگاری		مجموعه اولیه	
		RMSE	R-Sq	RMSE	R-Sq	RMSE	R-Sq	RMSE	R-Sq	RMSE	R-Sq
1 (20)	%10 (2)	0.073	0.853	0.068	0.891	0.075	0.836	0.064	0.881	0.045	0.943
	%15 (3)	0.067	0.873	0.059	0.920	0.063	0.887	0.048	0.933		
	%20 (4)	0.066	0.877	0.060	0.908	0.049	0.932	0.042	0.948		
2 (40)	%10 (4)	0.055	0.916	0.065	0.886	0.061	0.892	0.063	0.886	0.047	0.939
	%15 (6)	0.052	0.924	0.057	0.917	0.057	0.907	0.050	0.934		
	%20 (8)	0.051	0.923	0.056	0.918	0.060	0.898	0.052	0.924		
3 (60)	%10 (6)	0.046	0.943	0.056	0.918	0.042	0.948	0.048	0.934	0.045	0.943
	%15 (9)	0.048	0.937	0.052	0.933	0.045	0.940	0.041	0.953		
	%20 (12)	0.049	0.931	0.053	0.933	0.045	0.946	0.045	0.945		
4 (80)	%10 (8)	0.051	0.930	0.067	0.898	0.052	0.929	0.048	0.935	0.045	0.943
	%15 (12)	0.053	0.925	0.060	0.920	0.046	0.943	0.047	0.942		
	%20 (16)	0.051	0.927	0.057	0.922	0.050	0.933	0.043	0.951		
5 (100)	%10 (10)	0.063	0.893	0.068	0.888	0.055	0.924	0.045	0.948	0.047	0.939
	%15 (15)	0.056	0.917	0.066	0.895	0.049	0.939	0.045	0.947		
	%20 (20)	0.059	0.905	0.063	0.903	0.051	0.934	0.043	0.952		



۲ مقایسه درصد اشباع پیش‌بینی شده توسط مجموعه یادگیری اولیه (۲۰ عضوی) با مجموعه هرس شده (۲۰ درصد) بر مبنای داده‌های هدف (مغزه) شکل

ترسیم نمودار پراکنش بین ورودی‌های مختلف با خروجی که داده حاصل از مغزه بوده است، تعیین گردیده‌اند.

#### ۵- نتایج و بحث

جدول ۲- نتایج حاصل از اعمال متد پیشنهادی روی داده‌های چاه‌پیمایی اشاره شده در بخش قبلی را نشان می‌دهد. اولین ستون سمت چپ این جدول، تعداد خبره‌های مجموعه یادگیری اولیه را نمایش می‌دهد که به مجموعه‌های ۲۰، ۴۰، ۶۰، ۸۰ و ۱۰۰ عضوی تقسیم شده‌اند. ستون دوم از سمت چپ نیز معرف تعداد عناصر مجموعه هرس شده است که سه دسته ۱۰، ۱۵ و ۲۰ درصدی از عناصر مجموعه اصلی را تشکیل می‌دهد. چهار زوج ستون بعدی نیز به ترتیب عملکرد زیر مجموعه‌های هرس شده، تحت متدهای همبستگی، کو واریانس، کای-دو و ناسازگاری را در مقایسه با مجموعه یادگیری اصلی (آخرین زوج ستون سمت راست) برای سه دسته مذکور در هر بخش نمایش می‌دهد. در واقع، در هر بخش، یک مجموعه اصلی داریم که از صد درصد خبره‌های اولیه تشکیل شده و سه زیرمجموعه هرس شده که به ترتیب از ۱۰، ۱۵ و ۲۰ درصد کل خبره‌ها تشکیل گردیده است. در هر یک از پنج بخش (سطرهای ۱ تا ۵)، یک زوج متر اندازه‌گیری عملکرد که با (RMSE, R-Sq) نمایش داده شده به صورت پررنگ آمده که معرف بهترین عملکرد در بخش مربوطه است و متدی که این نتیجه را ارائه کرده است، متد برنده نامیده می‌شود. بنابراین، در چهار دسته (۲۰، ۶۰، ۸۰ و ۱۰۰ تایی) از خبره‌ها، عملکرد مجموعه هرس شده با متد گوناگونی ناسازگاری از بقیه بهتر بوده و در واقع روش ناسازگاری، بهترین عملکرد را جهت تعیین خبره‌های با گوناگونی بالا داشته که سبب بهبود در عملکرد و دقت مجموعه نهایی گردیده است.

تنها در دسته ۴۰ تایی خبره‌ها نتایج عملکرد مجموعه یادگیری دسته‌جمعی با کلیه خبره‌ها برنده بوده است. البته زیرمجموعه هرس شده با متد ناسازگاری در رتبه دوم قرار داشته که عملکردی نزدیک به مجموعه برنده دارد. از طرفی، با توجه به کاهش حداقل ۸۰ درصدی در تعداد اعضای مجموعه هرس شده در مقایسه با مجموعه‌های اصلی، استفاده از متد پیشنهادی (با توجه به کاهش در فضای حافظه‌ی مورد نیاز و سرعت عملیات محاسباتی)، بسیار مقرون به صرفه‌تر است. شکل ۲- همچنین، نمونه‌ای از مقایسه بین درصد اشباع پیش‌بینی شده توسط مجموعه یادگیری اولیه با مجموعه هرس شده بر مبنای داده‌های مغزه را برای مجموعه اولیه ۲۰ عضوی نمایش می‌دهد. مجموعه هرس شده نشان داده شده در شکل، همان مجموعه برنده این بخش است که تحت

برای اجرای هر یک از دو متد فوق، ابتدا باید اندازه زیرمجموعه نهایی (هرس شده) مشخص گردد. برای تعیین تعداد اعضای زیرمجموعه هرس شده، دو روش عمده وجود دارد که اولی، مقدار ثابت از پیش تعیین شده یا درصدی از تعداد کل است و دومی، به صورت پویا بوده که به میزان دقت زیرمجموعه نهایی با اندازه‌های مختلف وابسته است. سپس، با الگو گرفتن از متد پیشنهادی ژو (۲۰۰۲)، تعداد خبره‌هایی که دارای وزن بیشتر از یک مقدار مشخص هستند را شمرده و عدد حاصل (مثلاً  $\lambda$ ) را به عنوان اندازه زیرمجموعه نهایی تعیین می‌کنیم [۳۶]. پس از اینکه اندازه زیرمجموعه هرس شده مشخص شد، باید اعضای این مجموعه نیز انتخاب گردند. بدین منظور، باید به تعداد  $\frac{\lambda}{2}$  زوج خبره‌هایی که دارای عدد گوناگونی بیشتری هستند را با استفاده از معادلات ۱ تا ۴ انتخاب و به عنوان اعضای زیرمجموعه نهایی برگزید [۳۴].

#### ۴- روش تحقیق

در این مقاله، اثر انتخاب پیش‌بینی کننده‌های با گوناگونی بالا را روی دقت کل گروه، مورد بررسی و مطالعه قرار می‌دهیم. در ابتدا به تولید تعداد ۱۰۰ خبره از نوع شبکه عصبی با متد یادگیری LM می‌پردازیم. تفاوت این خبره‌ها در دستکاری انجام شده در تعداد داده‌های یادگیری ورودی است که از طریق نمونه‌گیری خود راه‌انداز انجام می‌پذیرد. تعداد خبره‌ها در گروه یادگیری اصلی اولیه، به پنج مجموعه ۲۰، ۴۰، ۶۰، ۸۰ و ۱۰۰ عضوی تقسیم می‌گردد. تعداد اعضای مجموعه هرس شده برای هر یک از گروه‌های اصلی ذکر شده، سه دسته ۱۰، ۱۵ و ۲۰ درصدی از کل هر یک از مجموعه‌ها را تشکیل می‌دهد. عملکرد مجموعه هرس شده، تحت چهار مدل انتخاب خبره‌های گوناگون در هر مرحله با مجموعه‌های اصلی مربوطه مقایسه شده است (جدول ۲). در ادامه، برای ترکیب نتیجه خروجی خبره‌ها از روش میانگین‌گیری ساده استفاده شده است. البته باید توجه داشت که این، تنها روش ترکیبی برای به دست آوردن نتیجه پیش‌بینی نهایی یک مجموعه یادگیری تجمعی نیست و می‌توان از روش‌های دیگری نظیر میانگین‌گیری موزون نیز در این مرحله استفاده کرد. متر استفاده شده برای مقایسه عملکرد هر یک از مدل‌ها، مجذور میانگین مربعات خطا و ضریب همبستگی است. نهایتاً، روش پیشنهادی روی داده‌های استفاده شده در مقاله [۱۸] که پیش‌بینی اشباع آب با استفاده از داده‌های چاه‌پیمایی بوده، اعمال گردید که نتایج آن در بخش بعدی آمده است. داده‌های چاه‌پیمایی به کار گرفته شده جهت ورودی شبکه عصبی شامل نمودارهای مقاومت الکتریکی، اشعه گاما، نوترون، چگالی و فاکتور فتوالکتریک بوده که از طریق

خیلی متفاوت باشند، آنگاه ممکن است خطای پیش‌بینی بالا باشد. در این مقاله، ابتدا چند متر برای اندازه‌گیری گوناگونی خبره‌های یک مجموعه معرفی شد. سپس، روش دوبه‌دو جهت انتخاب خبره‌های با گوناگونی بیشتر به کار برده شد. مترهای معرفی شده عبارت بودند از: همبستگی، کوواریانس، کای-دو و ناسازگاری. سپس، نتایج اعمال متد پیشنهادی بر روی داده‌های چاه‌پیمایی جهت تعیین اشباع آب در یک مخزن نفت ارائه گردید. نتایج نشان می‌دهد که گروه نسبتاً کوچک ایجاد شده (۱۰-۲۰ درصد از اعضای کل مجموعه) با استفاده از معیار ناسازگاری بین خبره‌ها، عملکرد قابل قبولی در مقایسه با گروه‌های اصلی بزرگتر دارند. ■

روش اندازه‌گیری گوناگونی با متد ناسازگاری و با ۲۰ درصد اعضای اولیه به دست آمد.

### نتیجه‌گیری

متدهای یادگیری دسته‌جمعی به‌طور وسیعی برای بهبود دقت مدل‌های پیش‌بینی کننده، مورد استفاده قرار گرفته است. در این راستا، دو عامل، نقش تعیین‌کننده‌ای در طراحی مجموعه فوق ایفا می‌کند که یکی، دقت و دیگری، میزان گوناگون بودن خبره‌ها نسبت به هم است. اگر اغلب خبره‌های یک مجموعه دارای خطای پیش‌بینی پایین باشند، یعنی بسیار مشابه هم رفتار می‌کنند و در عوض، اگر

### پانویس‌ها

- |                              |                            |                             |
|------------------------------|----------------------------|-----------------------------|
| 1. Artificial neural network | 11. Diversity              | 21. Weighted averaging      |
| 2. Fuzzy theory              | 12. Homogeneous ensemble   | 22. Pruning                 |
| 3. Neural fuzzy              | 13. Heterogeneous ensemble | 23. Functional Network      |
| 4. Genetic algorithm         | 14. Sampling               | 24. Correlation coefficient |
| 5. Accuracy                  | 15. Resampling             | 25. Chi-square              |
| 6. Performance               | 16. Manipulating           | 26. Disagreement Measure    |
| 7. Train                     | 17. Independently          | 27. Pairwise                |
| 8. Test data                 | 18. Sequentially           | 28. Levenberg Marquardt     |
| 9. Ensemble                  | 19. Simultaneously         | 29. Bootstrap sampling      |
| 10. Committee Machine        | 20. Simple averaging       | 30. Root mean square error  |

### منابع

- [1] Helle.H.B and Bhatt.A, "Fluid saturation from well logs using committee neural networks", Petroleum Geoscience, vol. 8, pp. 109-118, 2002
- [2] Bhatt.A and Helle. H.B, "Committee neural networks for porosity and permeability prediction from well logs", Geophysical Prospecting, vol. 50, pp. 645-660, 2002.
- [3] Nilsson.N.J, "Learning machines; foundations of trainable pattern-classifying systems", New York: McGraw-Hill, 1965.
- [4] Chen. C.-H and Lin. Z.-S, "A committee machine with empirical formulas for permeability prediction," Computers & Geosciences, vol. 32, pp. 485-496, 2006.
- [5] Kadkhodaie-Ilkhchi. A, Rezaee. M.R, and Rahimpour-Bonab. H, "A committee neural network for prediction of normalized oil content from well log data: An example from South Pars Gas Field, Persian Gulf," Journal of Petroleum Science and Engineering, vol. 65, pp. 23-32, 2009c.
- [6] Wong. P.M, Jang. M, Cho. S, and Gedeon.T. D, "Multiple permeability predictions using an observational learning algorithm," Comput. Geosci., vol. 26, pp. 907-913, 2000.
- [7] Kadkhodaie-Ilkhchi. A, Rezaee.M.R, Rahimpour-Bonab. H, and Chehrizi. A, "Petrophysical data prediction from seismic attributes using committee fuzzy inference system", Comput. Geosci., vol. 35, pp. 2314-2330, 2009.
- [8] Labani. M. M, Kadkhodaie-Ilkhchi. A, and Salahshoor. K, "Estimation of NMR log parameters from conventional well log data using a committee machine with intelligent systems: A case study from the Iranian part of the South Pars gas field, Persian Gulf Basin", Journal of Petroleum Science and Engineering, vol. 72, pp. 175-185, 2010.
- [9] Rezaee.M.R, Kadkhodaie Ilkhchi. A, and Barabadi. A, "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, vol. 55, pp. 201-212, 2007.
- [10] Ghiasi-Freez. J, Kadkhodaie-Ilkhchi. A, and Ziiai. M, "Improving the accuracy of flow units prediction through two committee machine models: An example from the South Pars Gas Field, Persian Gulf Basin, Iran", Computers & Geosciences, vol.



- 46, pp. 10-23, 2012.
- [11] Chandra. A and Yao. X, "DIVACE: Diverse and Accurate Ensemble Learning Algorithm," in *Intelligent Data Engineering and Automated Learning* vol. 3177, 2004, pp. 619-625.
- [12] Brown. G, Wyatt. J, Harris. R, and Yao. X, "Diversity creation methods: a survey and categorisation", 2005.
- [13] L. K. Hansen and P. Salamon, "Neural Network Ensembles," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, pp. 993-1001, 1990.
- [14] Krogh. A and Vedelsby. J, "Neural network ensembles, cross validation, and active learning", *Advances in Neural Information Processing Systems*, vol. 7, pp. 231-238, 1995.
- [15] Chandra. A and Yao. X, "Evolving hybrid ensembles of learning machines for better generalisation," *Neurocomput.*, vol. 69, pp. 686-700, 2006a.
- [16] Jafari Kenari. S.A, Mashohor. S, and Jalali Varnamkhasti. M, "Committee neural networks with fuzzy genetic algorithm," *Journal of Petroleum Science and Engineering*, vol. 76, pp. 217-223, 2011.
- [17] Asoodeh. M and Bagheripour. P, "Prediction of Compressional, Shear, and Stoneley Wave Velocities from Conventional Well Log Data Using a Committee Machine with Intelligent Systems", *Rock Mechanics and Rock Engineering*, pp. 1-19, 2011.
- [18] Jafari Kenari. S.A and Mashohor. S, "Robust committee machine for water saturation prediction", *Journal of Petroleum Science and Engineering*, vol. 104, pp. 1-10, 2013.
- [19] Sharkey. A. J. C, "On Combining Artificial Neural Nets", *Connection Science*, vol. 8, pp. 299 - 314, 1996.
- [20] Brown. G, Wyatt. J.L, and Tino. P, "Managing Diversity in Regression Ensembles", *Machine Learning Research*, vol. 6, pp. 1621-1650, 2005.
- [21] Partalas. I, Tsoumakas. G, and Vlahavas. I, "An ensemble uncertainty aware measure for directed hill climbing ensemble pruning", *Machine Learning*, vol. 81, pp. 257-282, 2010.
- [22] Zhang. Y, Burer. S, and Street. W.N, "Ensemble Pruning Via Semi-definite Programming", *Journal of Machine Learning Research*, vol. 7, pp. 1315-1338, 2006.
- [23] Hernandez-Lobato. D, Martinez-Munoz. G, and Suarez. A, "Pruning in Ordered Regression Bagging Ensembles", *Proceedings of the IEEE World Congress on Computational Intelligence*, 2006b, pp. 266-1273.
- [24] Martinez-Munoz. G and Suarez. A, "Pruning in ordered bagging ensembles", presented at the Proceedings of the 23rd international conference on Machine learning, 2006.
- [25] Martinez-Muoz. G, Hernandez-Lobato. D, and Suarez. A, "An Analysis of Ensemble Pruning Techniques Based on Ordered Aggregation", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, pp. 245-259, 2009.
- [26] Hernández-Lobato. D, Hernández-Lobato. J, Ruiz-Torrubiano. R, and Valle. A, "Pruning Adaptive Boosting Ensembles by Means of a Genetic Algorithm", *Intelligent Data Engineering and Automated Learning* ed, 2006a, pp. 322-329.
- [27] Hernández-Lobato. D, Martínez-Muñoz. G, and Suárez. A, "Empirical analysis and evaluation of approximate techniques for pruning regression bagging ensembles", *Neurocomputing*, vol. 74, pp. 2250-2264, 2011.
- [28] Bian. S and Wang. W, "On diversity and accuracy of homogeneous and heterogeneous ensembles", *International Journal of Hybrid Intelligent Systems*, vol. 4, pp. 103-128, 2007.
- [29] Shun. B and Wenjia. W, "Investigation on Diversity in Homogeneous and Heterogeneous Ensembles", *International Joint Conference on Neural Networks*, 2006, pp. 3078-3085.
- [30] Al-Bulushi, N., King, P. R., Blunt, M. J., and Kraaijeveld, M., "Development of artificial neural network models for predicting water saturation and fluid distribution", *Journal of Petroleum Science and Engineering*, vol. 68, pp. 197-208, 2009.
- [31] Adeniran, A., Elshafei, M., and Hamada, G., "Functional Network softsensor for formation porosity and water saturation in oil wells" in *International Instrumentation and Measurement Technology Conference*, pp. 1138-1143, 2009.
- [32] Alimoradi, A., Moradzadeh, A., and Bakhtiari, M. R., "Methods of water saturation estimation: Historical perspective" *Journal of Petroleum and Gas Engineering* vol. 2 (3), pp. 45-53, 2011.
- [33] Kamalyar, K., Sheikhi, Y., Jamialahmadi, M., 2012. Using an artificial neural network for predicting water saturation in an Iranian oil reservoir. *Pet. Sci. Technol.* 30(1), 35-45.
- [34] Dutta. H, "Measuring diversity in regression ensembles", *IJCAI*, pp. 2220-2236, 2009.
- [35] Tin Kam. H, "The random subspace method for constructing decision forests", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, pp. 832-844, 1998.
- [36] Zhou. Z.-H, Wu. , and Tang. W, "Ensembling neural networks: Many could be better than all", *Artificial Intelligence*, vol. 137, pp. 239-263, 2002.