

برآورد تخلخل سنگ مخزن با استفاده از ترکیب آنسامبلی خطی شبکه‌های عصبی مصنوعی منفرد براساس روش‌های تحلیلی و الگوریتم ژنتیک

محمود ذاکری^{۱*} و ابوالقاسم کامکار روحانی^۲

^۱ دانش‌آموخته کارشناسی ارشد ژئوفیزیک، دانشکده مهندسی معدن، نفت و ژئوفیزیک، دانشگاه صنعتی شاهرود، ایران
^۲ استادیار، دانشکده مهندسی معدن، نفت و ژئوفیزیک، دانشگاه صنعتی شاهرود، ایران

(دریافت: ۸۶/۶/۲۰، پذیرش نهایی: ۸۸/۱۲/۱۸)

چکیده

ترکیب آنسامبلی شبکه‌های عصبی مصنوعی نوعی از ماشین کمیت‌های با ساختار موازی است که خروجی شبکه‌های منفرد را با هدف تلفیق اطلاعات حاصل از هریک از آنها و در نتیجه دست‌یابی به نتیجه بهتر ترکیب می‌کند. در این تحقیق از ترکیب آنسامبلی شبکه‌های عصبی مصنوعی منفرد به منظور برآورد تخلخل مؤثر سنگ مخزن گازی کنگان در میدان عظیم هیدروکربوری پارس جنوبی استفاده شده است. داده‌های چاه‌نگاری ۴ چاه این میدان در بازه عمقی سازند کنگان مورد استفاده قرار گرفت. داده‌های نگارهای صوتی، چگالی، پرتو گاما و تخلخل نوترونی درحکم ورودی شبکه‌ها و تخلخل مؤثر به‌منزله خروجی شبکه‌ها در نظر گرفته شد. ابتدا شبکه‌های عصبی منفرد پس‌انتشار خطا با ساختارهای گوناگون به روش فرآیند آموزش طراحی شد و سپس ۷ مورد از آنها که بهترین نتایج، یعنی کمترین میانگین مربعات خطای برآورد تخلخل را برای الگوهای آزمون داشتند به‌منظور ساخت ترکیبات آنسامبلی انتخاب شدند. ترکیبات آنسامبلی خطی به روش میانگین‌گیری ساده، روش‌های ترکیب خطی بهینه هاشم و روش غیرتحلیلی ترکیب خطی بهینه با استفاده از الگوریتم ژنتیک ساخته شد. بهترین ترکیب آنسامبلی حاصل که بیشترین کاهش را در میانگین مربعات خطای برآورد تخلخل الگوهای آزمون نسبت به بهترین شبکه عصبی منفرد ایجاد کرده است، ترکیب خطی بهینه‌ای با استفاده از روش بهینه‌یابی الگوریتم ژنتیک بود که میانگین مربعات خطای برآورد تخلخل الگوهای آموزش و آزمون را به ترتیب ۱۴/۴٪ و ۱۲/۵٪ نسبت به بهترین شبکه عصبی منفرد کاهش داد.

واژه‌های کلیدی: تخلخل، نگار، شبکه عصبی مصنوعی، ماشین کمیت‌های، ترکیب آنسامبلی، الگوریتم ژنتیک.

Estimation of reservoir rock porosity using linear ensemble combination of single artificial neural networks based on analytical and genetic algorithm techniques

Zakeri, M.¹ and Kamkar-Rouhani, A.²

¹ M.Sc. graduate in Geophysics, School of Mining, Petroleum and Geophysics, Shahroud University of Technology, Iran

² Assistant Professor of School of Mining, Petroleum and Geophysics, Shahroud University of Technology, Iran

(Received: 11 Sep 2007, Accepted: 9 March 2010)

Abstract

Porosity is one of the most important properties for comprehensive studies of hydrocarbon reservoirs. For determination of porosity in a rock, that is the ratio of volume of voids to the total volume of the rock, there are two conventional methods: In the first method, direct measurement of porosity is carried out by testing drilling cores. In the second method, porosity is determined indirectly using well logging data and relevant

mathematical relations or equations. There are some limitations and difficulties for determination of porosity using both the above methods. Using the artificial neural networks (ANNs) method for this purpose can reduce these difficulties remarkably, and also, contains acceptable results. Solving any problem using ANNs needs a three-step procedure: training, generalization and operation. In the training step, the network teaches the patterns that exist in the inputs and the relation between the inputs and the outputs of the training set. Generalization is the ability of the network to present acceptable responses for the inputs that have not been included in the training set (unseen patterns). Operation is the use of the network for the objective problem. Obviously the network, which is used in the operation step, must be well trained and have a suitable generalization performance. One of difficulties which may occur for a network after being trained, is overfitting that is the same as poor generalization performance. If conditions are so that the network is trained to a favorable amount of error reduction for training patterns or to a distinct number of epochs but overfitting does not occur, in this state the training is called overtraining. In the ANNs method, a number of networks are trained. These networks are evaluated using a suitable performance criterion, for example mean square errors (MSE), and based on this criterion, the best network is selected. Although selecting the best single neural network generates the best obtained pattern, it leads to loss of information existing in the other networks. There is the drawback that the generalization performance of the best selected network for unseen patterns is limited and more over, error in estimation is common. If we accept that for all possible test patterns, complete or 100 % generalization is impossible, we have a convincing reason to search for methods for improving the performance of ANNs. For this purpose, a combination of trained networks using suitable methods has been proposed because this work may lead to integrate the information of the networks of the components in the combination and thus to help the enhancement of the accuracy of the results and the generalization performance of the combination in comparison with the best selected network. Using a combination of single neural networks, multiple network systems which are also called committee machine (CM), are generated to access better results for problems that a network alone cannot solve or may be solved effectively using CM. Ensemble combination of ANNs is a type of CM having parallel structure in which any of its components or networks solely presents a solution for the objective problem, and then the solution results are combined in a proper manner. In function estimation problems, ensemble combinations can be made linearly or nonlinearly. In this research work, linear ensemble combination of single artificial neural networks was applied in order to estimate the effective porosity of the Kangan gas reservoir rock in the giant Southern Pars hydrocarbon field. From the view point of structural geology, the Kangan gas deposit is an asymmetric anticline with a northwest-southeast spread whose southeast side is turned. This geologic formation consists of dolomite, limestone, dolomitic lime and thin layers of shale. Well logging data acquired from 4 wells in the area at a depth interval corresponding to the Kangan formation were used. 215 selected patterns from wells SP1, SP3 and SP13 were used for training the networks and 89 selected patterns from well SP6 were used for testing the generalization performance of the networks. In each pattern, acoustic, density, gamma ray and neutron porosity well log data were considered as the inputs of the networks and the effective porosity data were assigned as the outputs of the networks. First, back propagation single neural networks having different structures (totally 90 structures) were trained using the overtraining method. Then, 7 networks which had the best results, i.e. containing minimum MSE in the test step, were selected for making ensemble combinations. 120 Linear ensemble combinations of these 7 networks (i.e. 21 two-fold combinations, 35 three-fold combinations, 35 four-fold combinations, 21 five-fold

combinations, 7 six-fold combinations and 1 seven-fold combination) were constructed using analytical methods including simple averaging and four different Hashem's optimal linear combination (OLC) methods, i.e. unconstrained MSE-OLC with a constant term, constrained MSE-OLC with a constant term, unconstrained MSE-OLC without a constant term and constrained MSE-OLC without a constant term. In Hashem's methods, coefficients of networks in MSE-OLC are computed by performing a set of matrix operations. Then the best produced combination using the above-mentioned 5 analytical methods was selected from each of two-fold, three-fold, four-fold, five-fold, six-fold and seven-fold combination sets (i.e. the combination which contained minimum MSE in the test step). For the 6 selected combinations, in addition to analytical methods, the coefficients of MSE-OLC were computed using genetic algorithm (GA). The best produced analytical ensemble combination, which with respect to the other analytical combinations had the maximum reduction in MSE of the test step compared to the best single neural network, was a three-fold unconstrained MSE-OLC without constant term. This combination in comparison with the best single neural network decreased the MSE in the training and test steps 6.3 % and 4.9 %, respectively. Despite this, the best ensemble combination among all the combinations was a six-fold OLC obtained using the GA optimization method. This best ensemble combination, compared to the best single neural network, reduced the MSE in the training and test steps 14.4% and 12.5%, respectively. Generally, in the all cases that were investigated, OLC using GA yielded better results as it caused more reduction in MSE of the test step compared to analytical combinations. However, OLCs using Hashem's methods compared to other combinations generally contained more reductions in MSE of the training step.

Key words: Porosity, Log, Artificial neural network, Committee machine, Ensemble combination, Genetic algorithm.

۱ مقدمه

با آزمایش روی مغزه‌ها است. در ارتباط با اندازه‌گیری‌های مغزه، مسائل و مشکلاتی وجود دارد. مغزه‌گیری و به‌دست آوردن اطلاعات از مغزه مشکل، وقت‌گیر و بسیار پرهزینه است و به همین دلیل محاسبه تخلخل از روی مغزه در همه‌جاها کاری غیرمعمول است. تغییر شرایط مؤثر روی مغزه در آزمایشگاه نسبت به عمق برداشت و در نتیجه تغییر در خصوصیات فیزیکی آن باید مورد توجه قرار گیرد. همچنین ممکن است خواص مغزه کوچک برداشته شده در یک عمق معین، نشان‌دهنده خواص حجم کافی از سازند در این عمق نباشد. علاوه‌براین، اندازه‌گیری‌های آزمایشگاهی روی مغزه، در معرض خطا قرار دارد. روش دوم که غیرمستقیم است استفاده از نگارهای چاه است. علاوه بر سهولت و مزایای

تخلخل هر سنگ نسبت حجم فضاهای خالی موجود در آن به حجم کلی سنگ است. تخلخل مؤثر، تخلخل پیوسته‌ای است که در آن امکان حرکت شاره و در نتیجه دسترسی و استحصال آن وجود دارد و بنابراین تخلخل ناپیوسته، خلل و فرج اشغال شده با آب جذبی و آب محصور در شیل‌ها را شامل نمی‌شود. در واقع این نوع تخلخل است که مورد نظر مهندسان و کارشناسان نفتی است زیرا تراوایی سنگ و توان بهره‌دهی آن به میزان تخلخل مؤثر بستگی دارد.

تعیین تخلخل در حکم یکی از پارامترهای بسیار مهم و اساسی در تحقیقات جامع مخزن، از اهمیت خاصی برخوردار است. به‌طور معمول برای تعیین تخلخل مخزن دو راه وجود دارد: روش اول اندازه‌گیری مستقیم تخلخل

اقتصادی روش‌هایی که تداخل را براساس اندازه‌گیری‌های نگارها تعیین می‌کنند، این روش‌ها اطلاعات را به صورت پیوسته در اختیار کارشناس قرار می‌دهند و در یک عمق مشخص، از حجم بزرگ‌تری از سازند تأثیر می‌پذیرند. نگاری وجود ندارد که مستقیماً بتواند تداخل را اندازه‌گیری کند اما متخصصان از روش‌های گوناگونی استفاده می‌کنند تا تداخل را از داده‌های نگارها محاسبه کنند (هرست و همکاران، ۲۰۰۰).

روش‌هایی که به‌طور معمول از آنها برای تعیین تداخل و یا هر ویژگی مخزنی دیگر از روی نگارهای چاه استفاده می‌شود شامل فرض‌ها و شرایطی هستند که در صورت اقتناع آنها، نتایج حاصل صحیح است و در غیر این صورت می‌باید تصحیحات لازم به روشی مناسب اعمال شود. همچنین روابط و معادلاتی که از آنها در این روش‌ها استفاده می‌شود شامل ضرایب و عبارتهایی هستند که به نقطه قرائت و سنگ‌شناسی و زمین‌شناسی آن بستگی دارد و به‌طور کلی ناشناخته‌اند و بایستی با تحلیل نمونه‌های سنگی تعیین شوند، مثلاً ناخالصی رسی، نوع شاره منافذ، چگالی دانه‌ای در روابط مربوط به نگار چگالی و زمان گذر موج در دانه‌ها در روابط مربوط به نگار صوتی.

چنانچه از روش شبکه عصبی مصنوعی (Artificial neural network یا به‌طور خلاصه ANN) برای برآورد تداخل استفاده شود، حداقل می‌توان گفت بعد از آنکه شبکه مناسب طراحی شد، در مرحله اجرا (Operation) و کاربرد، ما را از فرمول‌ها و روابطی که نیاز به تعیین ضرایب و عبارات لازم دارند و همچنین تصحیحات ناشی از محقق نشدن فرض‌ها و شرایط حاکم بر روش مورد استفاده بی‌نیاز می‌کند زیرا شبکه این موارد را در مرحله آموزش (Training) آموخته و در مرحله تعمیم (Generalization)، توانایی شبکه در اعمال آنها مورد آزمون قرار گرفته است.

تحقیقات زیادی در زمینه کاربرد شبکه‌های عصبی

مصنوعی در تحلیل کمی خصوصیات و ویژگی‌های مخزن با استفاده از داده‌های نگارها صورت گرفته است، مثلاً هوانگ و ویلیامسون (۱۹۹۷) و هوانگ و همکاران (۱۹۹۶) از روش شبکه عصبی به منظور برآورد تداخل و تراوایی در میدین هیدروکربوری فلات‌قاره شرق کانادا بهره گرفته‌اند. تحقیقات آنها نشان می‌دهد که روش شبکه عصبی نسبت به روش‌های رگرسیون خطی چندمتغیره، رگرسیون غیرخطی چندمتغیره و برآورد با استفاده از روابط تجربی و نیمه‌تجربی، نتایج بسیار بهتری دارد. هلی و همکاران (۲۰۰۱) نیز با استفاده از شبکه عصبی، تداخل و تراوایی را در یکی از مخازن دریای شمال با دقت کافی که بیشتر نیازهای عملی را جواب‌گو است ارزیابی کرده‌اند.

علی‌رغم همه توانایی‌ها و امتیازات روش شبکه عصبی مصنوعی، با توجه به اینکه برای همه حالت‌هایی که قبلاً شبکه عصبی آنها را تجربه نکرده است و به شبکه عرضه نشده‌اند برآورد ۱۰٪ درست غیرممکن است، بنابراین جست‌وجوی روشی برای افزایش دقت روش شبکه عصبی توجیه منطقی پیدا می‌کند. همچنان تلاش‌ها برای ارتقای این روش ادامه دارد. استفاده از سامانه‌های چندشبکه‌ای (Multiple networks systems) که از آنها به شبکه عصبی کمیته‌ای یا ماشین کمیته‌ای (Committee machine یا به‌طور خلاصه CM) نیز تعبیر می‌شود و می‌توان گفت ترکیبی از چندین شبکه‌اند که به‌طور منفرد و مجزا آموزش دیده‌اند، به‌این‌منظور پیشنهاد شده است. در رهیافت ماشین کمیته‌ای تأکید بر آن است که از همه اطلاعات مفید، حداکثر استفاده شود و نه اینکه با انتخاب یک شبکه به‌منزله بهترین شبکه، از اطلاعات مفید سایر شبکه‌ها صرف‌نظر شود.

پارامترو و همکاران (۱۹۹۶) روشن ساختند که استفاده از میانگین‌گیری ساده چندین شبکه با ساختار و مجموعه آموزشی یکسان و پارامترهای اولیه تصادفی متفاوت،

شبکه جزء بهتر است و به خوبی توانسته است تخلخل را برآورد کند. وی در برآورد تراوایی از روش مدوله‌ای استفاده کرده است. دامنه تراوایی به سه زیردامنه تقسیم و هریک از سه مدول دستگاه با استفاده از یکی از زیردامنه‌ها آموزش داده شده است. با هدف افزایش دقت، به جای استفاده از یک شبکه عصبی در هر مدول، از یک ترکیب آنسامبلی متشکل از ۵ شبکه عصبی منفرد در هر مدول استفاده کرده است. مقایسه نتایج روش شبکه عصبی منفرد با نتایج این ماشین کمیت‌های نشان می‌دهد که در برآورد تراوایی، ماشین کمیت‌های دقت بیشتر و خطای کمتری داشته است. برای محاسبه اشباع هریک از سه شاره مخزنی آب، نفت و گاز، از یک ماشین کمیت‌های استفاده شده که ترکیب آنسامبلی ۹ شبکه پرسپترون با ساختار ۱-۴-۴ است. هریک از این ترکیبات آنسامبلی یکی از سه بلوک ساختاری یک ترکیب مدوله‌ای را تشکیل می‌دهد. ترکیب آنسامبلی نتایج ۹ شبکه به خوبی توانسته است خطا را نسبت به شبکه‌های منفرد جزء کاهش دهد.

چن و لین (۲۰۰۶) نیز از یک ماشین کمیت‌های متشکل از سه فرمول تجربی برای برآورد تراوایی استفاده کرده‌اند. آنها از دو روش میانگین‌گیری ساده و میانگین‌گیری وزنی به منظور ترکیب فرمول‌ها بهره گرفته‌اند. در روش میانگین‌گیری وزنی، ضرایب بهینه فرمول‌ها با استفاده از روش الگوریتم ژنتیک تعیین شده است. نتایج نشان می‌دهد که میانگین مربعات خطای برآورد تراوایی الگوهای آموزش با استفاده از میانگین‌گیری وزنی ۶٪ نسبت به بهترین فرمول تجربی کاهش پیدا کرده است. همچنین میانگین مربعات خطای برآورد تراوایی الگوهای آزمون توان تعمیم، با استفاده از میانگین‌گیری وزنی و میانگین‌گیری ساده به ترتیب ۶٪ و ۱۵٪ نسبت به بهترین فرمول تجربی کاهش یافته است.

کفایت نداشتن و محدود بودن تحقیقات صورت گرفته

می‌تواند به مقدار قابل توجهی واریانس خطا را کاهش دهد و نسبت به روش انتخاب بهترین شبکه، کارایی بیشتری دارد. نافتالی و همکاران (۱۹۹۷) به بررسی عملکرد ماشین کمیت‌های پرداختند. بنابر نتایج آنها، ماشین کمیت‌های واریانس خطای برآورد را کاهش می‌دهد ولی کاهش در آریبی (bias) خطا ایجاد نمی‌کند، درحالی‌که هاشم (۱۹۹۷) روش ترکیب خطی بهینه (Optimal linear combination یا OLC) نامقید را به منظور حذف آریبی پیشنهاد کرد. شارکی و همکاران (۲۰۰۰) نیز رهیافت آزمون و انتخاب (Test and select approach) را به منظور انتخاب ترکیبات آنسامبلی (Ensemble combinations) کارآمد شرح داده‌اند.

میزان به کارگیری شبکه‌های عصبی کمیت‌های در علوم مربوط به زمین و از جمله چاه‌نگاری، نسبت به شبکه‌های عصبی منفرد، بسیار محدودتر است. چند مورد از آنها عبارت‌اند از: فانگ و همکاران (۱۹۹۷) از ترکیب مدوله‌ای شبکه‌های عصبی (Modular neural networks) یا به طور خلاصه (MNN) که نوعی از ماشین کمیت‌های است برای تشخیص جنس و درصد کانی‌های سنگ‌های مجاور یک چاه از روی نگارهای چاه استفاده کرده‌اند و نتایج را با شبکه پسانتشار خطای منفرد مورد مقایسه قرار دادند. مقایسه نشان می‌دهد که ترکیب مدوله‌ای شبکه‌های عصبی، مدت زمان آموزش و میانگین مربعات خطا (MSE) را نسبت به شبکه پسانتشار خطای منفرد به میزان زیادی کاهش داده است.

بات (۲۰۰۲) با استفاده از شبکه‌های عصبی کمیت‌های به برآورد تخلخل، تراوایی و اشباع شاره‌های مخزنی در میادین نفتی دریای شمال پرداخته است. وی برای برآورد تخلخل از ترکیب آنسامبلی شبکه‌ها که نوعی دیگر از ماشین کمیت‌های است بهره گرفته است. در ترکیب آنسامبلی مورد استفاده، نتایج ۹ شبکه منفرد با هم ترکیب شده است. نتیجه آنسامبل حاصل، از نتایج هریک از ۹

قسمت‌های گوناگون این نرون به شرح زیر است:

(۱) سیناپس‌ها که هر یک با وزنی مختص به خودش تعیین می‌شوند. سیگنال x_j در ورودی سیناپس λ_m با ضرب شدن در وزن w_{kj} با نرون k ارتباط پیدا می‌کند. (۲) یک جمع‌کننده که سیگنال‌های وزن‌دار را با هم جمع می‌کند. (۳) یک تابع انتقال یا فعال‌ساز که مقدار خروجی جمع‌کننده را به خروجی نرون تبدیل می‌کند. (۴) همچنین هر نرون یک آریبی (b_k) دارد که وظیفه آن کاهش و یا افزایش ورودی تابع انتقال است. به مجموعه وزن‌های سیناپسی و آریبی نرون‌ها، پارامترهای شبکه گفته می‌شود. خروجی نرون k به صورت زیر بیان می‌شود:

$$y_k = \varphi(v_k) = \varphi(u_k + b_k) \quad (1)$$

که در آن:

$$u_k = \sum_{j=1}^m w_{kj} x_j \quad (2)$$

برای حل هر مسئله‌ای، شبکه‌های عصبی سه مرحله را طی می‌کنند: آموزش، تعمیم و اجرا (حسنی پاک و شرف‌الدین، ۱۳۸۰). آموزش فرایندی است که طی آن، شبکه الگوی موجود در ورودی‌ها و ارتباط بین ورودی‌ها و خروجی‌های مجموعه الگوهای آموزشی را فرا می‌گیرد و منجر به دست‌یابی به پارامترهای بهینه برای هر نرون شبکه می‌شود. تعمیم، توانایی شبکه در عرضه جواب قابل قبول در گرو ورودی‌هایی که در مجموعه آموزشی نبوده‌اند است. اجرا، استفاده از شبکه در عملکردی است که به آن منظور طراحی شده است. بدیهی است که شبکه‌ای که در مرحله کاربرد و اجرا مورد استفاده قرار می‌گیرد بایستی به‌خوبی آموزش دیده باشد و از توان تعمیم مناسبی برخوردار باشد. یکی از مشکلاتی که در طی آموزش شبکه ممکن است پیش بیاید بیش‌برازش (Overfitting) و یا همان توانایی تعمیم کم است. اگر شرایط به گونه‌ای باشد که شبکه تا کاهش خطا به مقدار مطلوب برای الگوهای آموزش و یا به تعداد معینی دوره،

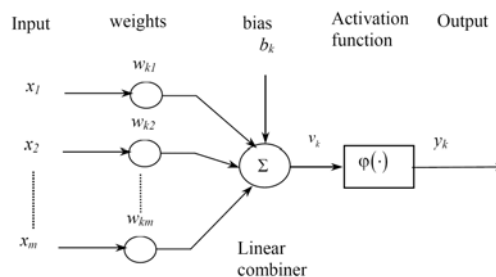
در زمینه ماشین کمیت‌های، به‌خصوص درباره کاربرد آن در تحقیقات مخزن، باعث شده است تا نقاط قوت و ضعف روش‌های متفاوت ساخت آنها مبهم باشد، به همین دلیل تحقیقات نظری و کاربردی بیشتر با هدف رفع این ابهامات ضروری است.

هدف از تحقیق حاضر، برآورد تخلخل سنگ مخزن با استفاده از دو رهیافت شبکه‌های عصبی منفرد و شبکه‌های عصبی کمیت‌های است تا در نهایت با مقایسه نتایج، ببینیم که آیا در این مورد تحقیقی خاص، ماشین کمیت‌های می‌تواند نتایج روش شبکه‌های عصبی منفرد را بهبود دهد یا خیر و در صورت مثبت بودن پاسخ، چه عواملی بر میزان بهبود و ارتقای نتایج تأثیر می‌گذارند.

متأسفانه در کشورمان تحقیقی در مورد کاربرد ماشین کمیت‌های در برآورد و تشخیص ویژگی‌ها و خصوصیات مخزنی مشاهده نشد و به نظر می‌رسد که این تحقیق گام اولی است که در این زمینه برداشته می‌شود. امیدواریم با بررسی در این مورد، مقداری از این خلأ را اگر چه بسیار اندک، جبران کنیم و این تحقیق بتواند نقطه آغازی مفید و مناسب برای پژوهش‌های آینده باشد.

۲ شبکه‌های عصبی مصنوعی

شبکه‌های عصبی مصنوعی سامانه‌هایی پویا و غیرخطی‌اند که تعداد زیادی واحد پردازش به نام نرون (Neuron) مصنوعی را دربرمی‌گیرند و به‌طور کلی در سه نوع لایه ورودی، میانی و خروجی تعبیه شده‌اند. شکل ۱ نمایشی از یک نرون مصنوعی است.



شکل ۱. نمایشی از یک نرون مصنوعی (بات، ۲۰۰۲).

از آنها، الگوریتم متوقف می‌شود، مثلاً بیشینه دفعات تولید نسل را می‌توان تعیین کرد و یا زمانی که مقدار برازش یکی از کروموزم‌های جمعیت جاری از یک حد تعیین شده بیشتر باشد تولید نسل متوقف می‌شود.

۴ شبکه‌های عصبی کمیته‌ای

چنان‌که گفته شد در روش شبکه عصبی مصنوعی، تعدادی شبکه آموزش دیده تولید می‌شود. این شبکه‌ها با استفاده از یک معیار عملکرد مناسب (مثلاً میانگین مربعات خطا) سنجش می‌شوند و براین اساس بهترین شبکه انتخاب می‌شود. انتخاب بهترین شبکه منفرد (NN_b)، اگرچه بهترین الگوی حاصل را تولید می‌کند اما باعث از دست رفتن اطلاعاتی می‌شود که در دیگر شبکه‌ها وجود داشته است. بنابراین به جای اینکه یکی از شبکه‌ها به منزله بهترین شبکه انتخاب و از بقیه صرف نظر شود، ترکیب شبکه‌های آموزش دیده با استفاده از روشی مناسب پیشنهاد شده است، زیرا این کار ممکن است به تلفیق اطلاعات شبکه‌های جزء ترکیب و در نتیجه افزایش دقت و توان تعمیم کمک کند. با استفاده از ترکیب شبکه‌های منفرد، سامانه‌های چندشبکه‌ای تولید می‌شود تا در مسئله‌هایی که یک شبکه قادر به حل آنها نیست و یا ممکن است با استفاده از سامانه‌های چندشبکه‌ای به‌طور مؤثرتری حل شوند به نتیجه بهتری دست یافت. دو روش معمول برای ایجاد سامانه‌های چندشبکه‌ای وجود دارد: ۱) ترکیب آنسامبلی که در این مقاله از آن استفاده شده است و در بخش بعدی شرح داده می‌شود. ۲) ترکیب مدوله‌ای که در آن مسئله به تعدادی زیرمسئله تفکیک و هر زیرمسئله با یک شبکه عصبی یا آنسامبلی از آنها حل می‌شود و در نهایت با تلفیق پاسخ‌ها، کل مسئله حل می‌شود. این دو نوع ترکیب با هم ناسازگار نیستند به این مفهوم که یک سامانه چندشبکه‌ای می‌تواند شامل هر دو نوع ترکیب باشد. های کین (۱۹۹۹)، سامانه‌های چندشبکه‌ای را با عنوان

آموزش ببند ولی دچار بیش‌برازش نشود این روش آموزش را فراآموزش (Overtraining) می‌نامند.

۳ الگوریتم ژنتیک

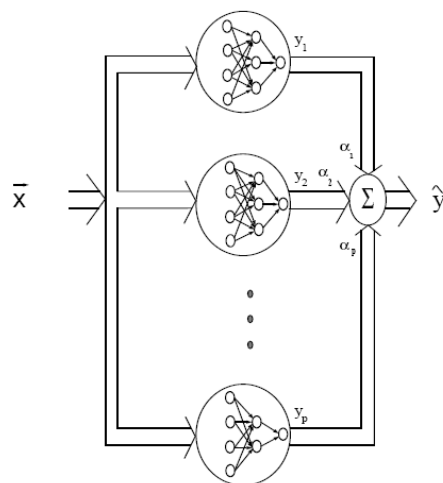
الگوریتم ژنتیک (Genetic Algorithm) یا به‌طور خلاصه GA) یک روش بهینه‌یابی قدرتمند است که براساس الگوی گزینش (Selection) و تکامل (Evolution) زیستی عمل می‌کند و می‌توان آن را در حل انواع گوناگونی از مسئله‌های بهینه‌یابی که با روش‌های متعارف به‌خوبی قابل حل نیستند به کار برد. این الگوریتم با ایجاد یک جمعیت (Population) اولیه از حل‌ها که در اصطلاح کروموزم (Chromosome) نامیده می‌شوند، شروع می‌شود. جمعیت جاری با هدف دست‌یابی به یک راه‌حل بهینه متحول می‌شود. به‌منظور ایجاد جمعیت جدید از جمعیت جاری، با استفاده از تابع خاصی، مقدار تابع هدف به ازای کروموزم‌های جمعیت جاری مقیاس‌بندی می‌شود تا مقدار برازش هریک از کروموزم‌ها تعیین شود. آن‌گاه با استفاده از روش‌هایی خاص و براساس مقدار برازش، تعدادی از کروموزم‌های جمعیت جاری به‌مثابه والد انتخاب می‌شوند. بنابر اصل نخبه‌گرایی (Elitism)، تعداد خاصی از کروموزم‌های جمعیت جاری که بیشترین برازش را داشته‌اند، بدون تغییر در نسل بعد حفظ می‌شوند و بقیه کروموزم‌های جمعیت جدید با استفاده از والدین و براساس دو قانون زیستی تقاطع (Crossover) و جهش (Mutation) تولید می‌شوند. در تقاطع، کروموزم‌های جدید که در اصطلاح به آنها بچه گفته می‌شود از ترکیب یک جفت از والدها تولید می‌شوند؛ ولی در جهش، بچه از اعمال تغییرات در یک والد ایجاد می‌شود. بعد از چندین بار تولید نسل، الگوریتم ژنتیک همگرا می‌شود و به یک راه‌حل بهینه تقریباً مطلق با برازش خوب برای تابع هدف می‌رسد. در الگوریتم ژنتیک، معیارهای گوناگونی برای توقف روند تولید نسل وجود دارد که با تحقق یکی

ماشین کمیته‌ای نام‌گذاری کرده است، هر چند که چنین روش‌هایی قبلاً شناخته شده و به کار رفته بود.

۵ ترکیب آنسامبلی

در یک آنسامبل، هریک از شبکه‌های جزء به تنهایی حلی را برای مسئله مورد نظر عرضه می‌کنند و نتایج به روشی مناسب با هم ترکیب می‌شود (بات، ۲۰۰۲). واضح است که ترکیب شبکه‌های یکسان فایده‌ای ندارد زیرا به‌طور یکسان تعمیم می‌یابند. تأکید می‌کنیم که شبکه‌ها بایستی به‌طور متفاوت و یا حداقل مشابه و نه یکسان تعمیم یابند. بنابراین به منظور ایجاد شبکه‌های سازنده ترکیب که این شرط را دارا باشند می‌توانیم یکی از موارد زیر را تغییر دهیم: پارامترهای اولیه شبکه، الگوهای آموزشی، ساختار شبکه‌ها و یا الگوریتم آموزش.

وقتی که مجموعه‌ای از شبکه‌ها برای ساخت ترکیبات آنسامبلی ایجاد شدند بایستی نتایج این شبکه‌ها با روشی مناسب و مؤثر با هم ترکیب شود. چندین روش متفاوت ترکیب وجود دارد که انتخاب یکی از آنها به نوع مسئله‌ای که می‌خواهیم آن را حل کنیم بستگی دارد. در مسئله‌های برآورد تابع، ترکیب آنسامبلی می‌تواند به‌طور خطی یا غیر خطی اجرا شود. در شکل ۲ نمایی از ترکیب خطی شبکه‌ها نشان داده شده است.



شکل ۲. ترکیب خطی شبکه‌های آموزش‌دیده (هاشم و همکاران، ۱۹۹۴).

چنان که در شکل دیده می‌شود، ترکیب خطی خروجی شبکه‌های گوناگون، مشابه ایجاد یک شبکه بزرگ است که شبکه‌های سازنده آن به‌طور موازی با هم عمل می‌کنند و ضرایب ترکیب، همان وزن‌های سیناپسی لایه خروجی هستند. برای ورودی \vec{x} ، خروجی ترکیب، \vec{y} ، جمع وزنی خروجی‌های متناظر شبکه‌های جزء (y_j ها) است و α_j ها ضرایب مربوط به آنها است. اختلاف اصلی در این تشبیه آن است که در ترکیب، ضرایب ارتباطی (α_j ها) ثابت هستند ولی در شبکه عصبی بزرگ معادل، این ضرایب که نقش وزن‌های سیناپسی لایه آخر را دارند در طول آموزش تغییر می‌کنند. مزیت دیگری که این ترکیب نسبت به شبکه هم‌ارز خود دارد آن است که در شبکه بزرگ معادلش، پارامترهای بیشتری وجود دارد که لازم است در فرایند آموزش تعیین شوند. بنابراین زمان آموزش زیادتر است و از طرفی به خاطر افزایش پارامترها، احتمال بیش‌برازش بیشتر می‌شود (هاشم و همکاران، ۱۹۹۴).

ترکیب خطی خروجی‌های p شبکه به کاررفته در ساختار ترکیب، عبارت است از:

$$\tilde{y}(\vec{x}; \vec{\alpha}) = \sum_{j=1}^p \alpha_j y_j(\vec{x}) \quad (3)$$

که خطای آن به صورت زیر است:

$$\delta(\vec{x}; \vec{\alpha}) = r(\vec{x}) - \tilde{y}(\vec{x}; \vec{\alpha}) \quad (4)$$

در این رابطه، $r(\vec{x})$ مقدار مطلوب برای ورودی \vec{x} است.

بنابراین لازم است که مقادیر مناسبی برای α_j ها پیدا کنیم.

یک روش همان انتخاب بهترین شبکه (NN_b) است که ضریب متناظر آن (α_b) را برابر یک قرار دهیم و بقیه ضرایب را صفر در نظر می‌گیریم. این روش اگرچه ساده است ولی چنان که قبلاً گفته شد، اطلاعات مفید $(p-1)$ شبکه دیگر را از دست می‌دهد.

است و این الگوها با

$K = \{K_n : K_n = (\vec{x}_n, r(\vec{x}_n), \vec{y}(\vec{x}_n)), n = 1, \dots, k\}$
 نشان داده می‌شوند. به مجموعه الگوهای K ، سری الگوهای ترکیب گفته می‌شود. به طور معمول K همان مجموعه الگوهای آموزشی شبکه‌ها انتخاب می‌شود. علاوه بر روش تحلیلی هاشم، الگوریتم ژنتیک نیز می‌تواند در حکم یک روش هوش مصنوعی که بیشترین کاربرد را در بین روش‌های بهینه‌سازی دارد، به منظور تعیین ضرایب بهینه با استفاده از سری الگوهای ترکیب به کار برد.

به طور مشابه با شبکه‌های عصبی منفرد، توان تعمیم OLC عبارت است از کارایی آن برای الگوهایی که در سری ترکیب نبوده‌اند و در برآورد ضرایب بهینه نقشی نداشته‌اند.

لازم است یادآوری کنیم که در این مقاله به منظور برآورد تخلخل، از روش‌های متفاوت ترکیب آنسامبلی خطی شبکه‌های عصبی (روش‌های تحلیلی شامل میانگین‌گیری ساده و روش‌های چهارگانه ترکیب خطی بهینه هاشم و روش غیرتحلیلی ترکیب خطی بهینه با استفاده از الگوریتم ژنتیک) استفاده شده است.

۶ برآورد تخلخل سنگ مخزن گازی کنگان در میدان پارس جنوبی با استفاده از ترکیب آنسامبلی خطی شبکه‌های عصبی مصنوعی منفرد

به منظور بررسی قابلیت و توانایی شبکه‌های عصبی مصنوعی کمیته‌ای در بهبود نتایج حاصل از شبکه‌های عصبی مصنوعی منفرد، بررسی موردی در میدان عظیم هیدروکربوری پارس جنوبی در بخش‌های گازدار k-1 و k-2 سازند کنگان انجام شد و تخلخل مؤثر ابتدا با استفاده از شبکه‌های عصبی مصنوعی منفرد و سپس با استفاده از ترکیب آنسامبلی خطی نتایج آنها برآورد شد.

۱-۶ زمین‌شناسی سازند کنگان

روش دوم، میانگین‌گیری ساده است که به طور گسترده‌ای مورد استفاده قرار می‌گیرد. در این روش فرض می‌شود که همه شبکه‌ها به یک میزان مفیدند و دقت و توانایی یکسانی در حل مسئله دارند. با استفاده از این روش کاهش در آریبی ایجاد نمی‌شود ولی واریانس ممکن است کاهش یابد.

روش سوم، یافتن ضرایب بهینه (OLC) است به طوری که تابع عملکرد را که معمولاً MSE در نظر گرفته می‌شود کمینه کند. به این منظور، هاشم (۱۹۹۳) روشی تحلیلی را پیشنهاد کرده است. در این روش، به منظور تصحیح آریبی y_j ها، جمله دیگری نیز به معادله ۳ افزوده می‌شود. این جمله $\alpha_0 y_0(\vec{x})$ است که در آن $y_0(\vec{x}) = 1$. بنابراین خروجی ترکیب به صورت زیر در خواهد آمد:

$$\vec{y}(\vec{x}; \vec{\alpha}) = \sum_{j=0}^p \alpha_j y_j(\vec{x}) = \vec{\alpha}' \vec{y}(\vec{x}) \quad (5)$$

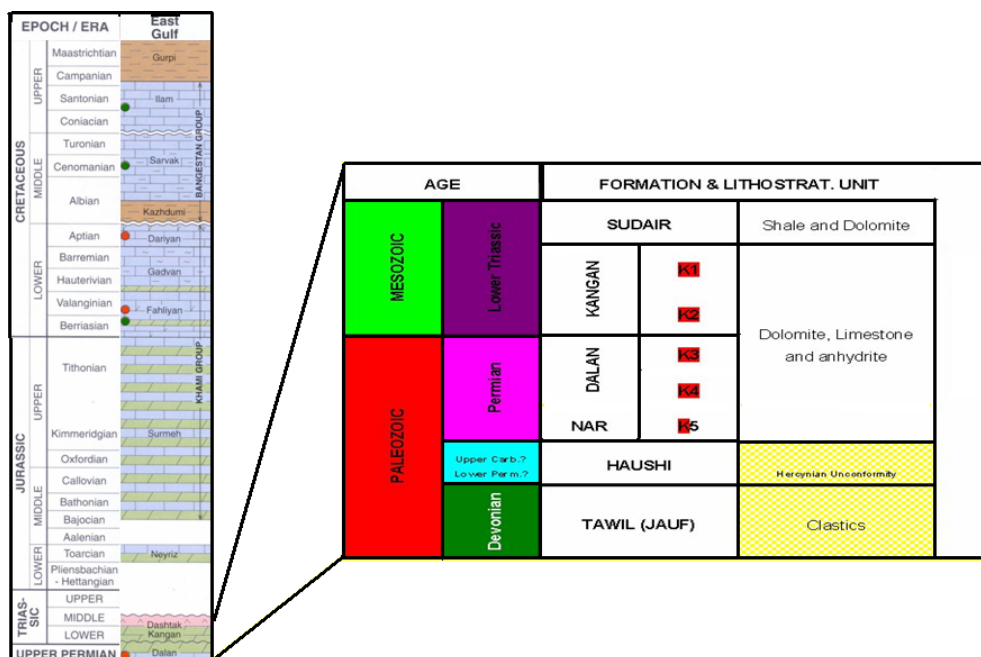
که $\vec{y}(\vec{x})$ و $\vec{\alpha}$ بردارهایی $(p+1) \times 1$ هستند. برای رابطه ۵ که حالت کلی را نشان می‌دهد سه حالت خاص نیز می‌توان در نظر گرفت. بنابراین در مجموع چهار حالت خواهیم داشت. تفاوت این چهار حالت به موجود بودن یا موجود نبودن جمله ثابت α_0 و شرط $\sum_{j=1}^p \alpha_j = 1$ مربوط می‌شود. اگر مجموع ضرایب شبکه‌ها ۱ باشد ترکیب را مقید (Constrained) و در غیر این صورت ترکیب را نامقید (Unconstrained) می‌نامند. پس این چهار حالت عبارت‌اند از ۱- نامقید با جمله ثابت (حالت کلی) ۲- مقید با جمله ثابت ۳- نامقید بدون جمله ثابت ۴- مقید بدون جمله ثابت.

ضرایب بهینه با اجرای یک سلسله عملیات ماتریسی و با استفاده از مجموعه‌ای از الگوهای تحقق‌یافته تعیین می‌شود که در آن به ازای ورودی \vec{x}_n ، مقدار تابع هدف و خروجی هر یک از شبکه‌های جزء آنسامبل مشخص

آلی و لایه های نازکی از شیل به رنگ های سبز و سبز مایل به خاکستری است. این بخش در چاه های متفاوت پارس جنوبی ۲۵ تا ۳۰ متر ضخامت دارد. بعد از این بخش، ناحیه گازدار سازند کنگان قرار دارد که به طور کلی از سنگ آهک و آهک دولومیتی سفید و سفید مایل به نخودی تا قهوه ای روشن بلورین شبه تخمکی متخلخل و تخریبی آلی تشکیل شده است. این قسمت دارای تخلخل مناسبی است و یکی از بخش های مهم گازدار میدان پارس جنوبی به شمار می رود. ضخامت این بخش در چاه های گوناگون به طور متوسط ۱۵۵ متر است (رحیمی، ۱۳۸۱). براساس توصیف مغزه و اطلاعات پتروفیزیکی و تلفیق تحقیقات میکروسکوپی و ماکروسکوپی و با به کار بردن خصوصیات از قبیل سنگ شناسی، رخساره و خصوصیات مخزنی و با توجه به تقسیم بندی توالی کربنات خوف (Khuff) در سپر عربستان، این ناحیه خود به دو واحد مخزنی مشخص و مجزا شامل k-1 و k-2 تقسیم شده است (زارعی و همکاران، ۱۳۸۶). در شکل ۳ ستون چینه شناسی میدان پارس جنوبی نمایش داده شده است.

میدان پارس جنوبی در موقعیت ۲۵ تا ۲۵/۵ درجه شرقی و ۲۶/۵ تا ۲۷ درجه شمالی و در ۱۰۵ کیلومتری جنوب غربی بندر عسلویه واقع شده و بزرگ ترین میدان دریایی در جهان است.

با توجه به اطلاعات به دست آمده از چاه های میدان پارس جنوبی و براساس نمونه های سنگی برداشت شده به هنگام حفاری و نمودارهای گرفته شده، رسوبات سازند کنگان متعلق به دوران تریاس زیرین گزارش شده است. این سازند با یک ناپیوستگی موازی روی سازند دالان به سن پرمین قرار گرفته و در بالا با شیل های سازند دشتک پوشیده شده است (خسروتهرانی، ۱۳۸۳). از نظر زمین شناسی ساختمانی، کانسارگازی کنگان تاقدیس نامتقارنی با امتداد شمال غربی - جنوب شرقی است که دامنه جنوب شرقی آن برگشته است (افشارحرب، ۱۳۸۱). با ظاهر شدن اولین لایه دولومیت نخودی مایل به خاکستری رس دار در زیر شیل های آغار، شروع سازند کنگان مشخص می شود. بخش فوقانی شامل دولومیت خاکستری - نخودی مایل به خاکستری - بلورین و تخریبی



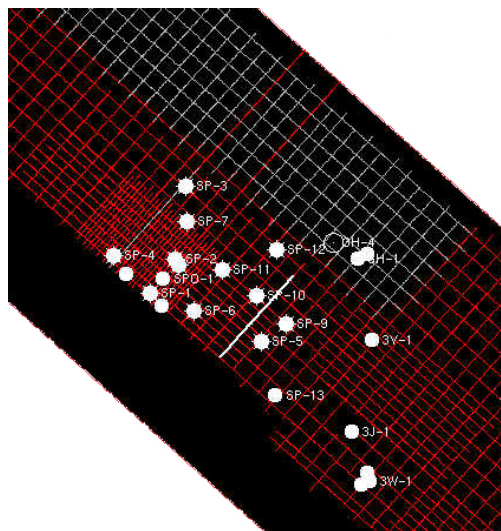
شکل ۳. ستون چینه شناسی میدان پارس جنوبی (رحیمی، ۱۳۸۱).

۲-۶ الگوهای آموزش و آزمون

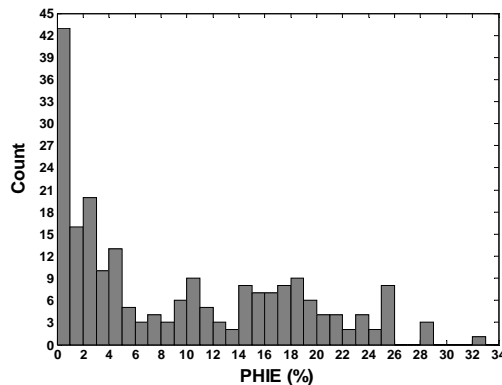
شکل ۴ موقعیت چاه‌های میدان پارس جنوبی را نشان می‌دهد. با توجه به محرمانه بودن موقعیت دقیق جغرافیایی چاه‌ها به لحاظ مشترک بودن این میدان بین ایران و قطر، مقیاس شکل و مختصات چاه‌ها در اختیار قرار نگرفت. داده‌های چاه‌نگاری چاه‌های SP1، SP3، SP6 و SP13 در بازه عمقی سازند کنگان (حداقل عمق، تقریباً ۲۶۰۰ متر در SP1 و حداکثر عمق، تقریباً ۳۰۰۰ متر در SP13) به منظور طراحی، آموزش و آزمون شبکه‌ها استفاده شد. از ۲۱۵ الگوی منتخب از چاه‌های SP1، SP3 و SP13 به منظور آموزش شبکه‌ها و از ۸۹ الگوی انتخاب شده از چاه SP6 برای آزمون توان تعمیم شبکه‌ها استفاده شد.

هر الگوی مورد استفاده، داده‌های نگارهای صوتی (DT)، چگالی (RHOB)، پرتو گاما (GR) و تخلخل نوترونی (NPHI) را درحکم ورودی و تخلخل مؤثر (PHIE) را درحکم خروجی شبکه‌ها در برمی‌گیرد. مقدار تخلخل مؤثر از روی داده‌های نگارها و با استفاده از تحقیقات مغزه و اطلاعات زمین‌شناسی عمقی پس از اعمال تصحیحات موردنیاز محاسبه شده است.

شکل ۵ نمودار توزیع فراوانی الگوها (Count) برحسب درصد تخلخل مؤثر ((PHIE(% را در الگوهای آموزشی نشان می‌دهد. با توجه به این شکل مشخص است که فراوانی الگوهای با تخلخل زیاد نسبت به الگوهای با تخلخل کم، کمتر است.



شکل ۴. موقعیت چاه‌های میدان پارس جنوبی (رحیمی، ۱۳۸۱).



شکل ۵. نمودار توزیع فراوانی مقادیر تخلخل مؤثر در ۲۱۵ الگوی آموزشی.

بیش‌برازش بوده است. در مورد هر ساختار خاص شبکه با نرون‌های میانی مشخص (مجموعاً ۹۰ ساختار)، یک مرحله آزمون و خطای طولانی با شرایط اولیه متفاوت صورت گرفت و بهترین نتیجه که کمترین میانگین مربعات خطای برآورد تخلخل را برای الگوهای آزمون داشت ثبت شد. سپس نتایج ساختارها براساس میانگین مربعات خطای برآورد تخلخل الگوهای آزمون مرتب شد و در نهایت از بین ۹۰ شبکه با ساختارهای متفاوت، هفت شبکه که کمترین میانگین مربعات خطای برآورد تخلخل الگوهای آزمون را داشتند برای ساخت ترکیبات آنسامبلی انتخاب شدند. در جدول ۱ نتایج این هفت شبکه آمده است. ستون‌های این جدول از سمت راست عبارت‌اند از شماره ردیف، ساختار شبکه، میانگین مربعات خطای برآورد شبکه برای الگوهای آموزش، ضریب همبستگی بین مقادیر تخلخل و برآورد شبکه برای الگوهای آموزش، میانگین مربعات خطای برآورد شبکه برای الگوهای آزمون، ضریب همبستگی بین مقادیر تخلخل و برآورد شبکه برای الگوهای آزمون و تعداد کل پارامترهای شبکه.

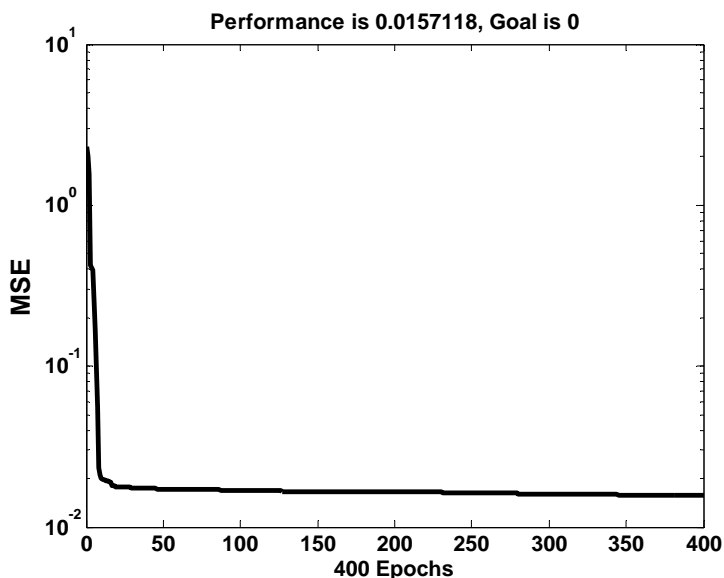
۳-۶ برآورد تخلخل با استفاده از شبکه‌های عصبی پس‌انتشار خطای آموزش‌دیده به روش فراآموزش و ترکیب آنسامبلی خطی نتایج آنها کلیه برنامه‌های مورد استفاده در این تحقیق و همچنین کارهای محاسباتی در محیط نرم‌افزار مطلب (MATLAB) تهیه و عملی شده است. در این تحقیق، ابتدا تخلخل با شبکه‌های منفرد پس‌انتشار خطای آموزش‌دیده به روش فراآموزش برآورد شده است و سپس نتایج شبکه‌های منفرد منتخب با هم ترکیب شده‌اند. به منظور پیش‌پردازش داده‌ها از تابع Prestd در محیط نرم‌افزار مطلب که مقادیر داده‌ها را به توزیع گاوسی استاندارد تبدیل می‌کند استفاده شد. برای آموزش از تابع آموزش پس‌انتشار خطای لونبرگ-مارکوارت (Trainlm) استفاده شد. تعداد دوره‌های تکرار برای آموزش شبکه‌ها ۴۰۰ مرتبه انتخاب شد. از شبکه‌هایی با یک و دو لایه میانی با تعداد نرون‌های متفاوت در این لایه‌ها استفاده شد. تابع انتقال نرون‌های لایه‌های میانی تابع تانزانت هذلولوی و تابع انتقال نرون‌های لایه خروجی تابع خطی انتخاب شد. معیار توقف روند افزایش نرون در لایه‌های میانی، افزایش احتمال

جدول ۱. ۷ شبکه منفرد آموزش‌دیده به روش فراآموزش با بهترین نتایج.

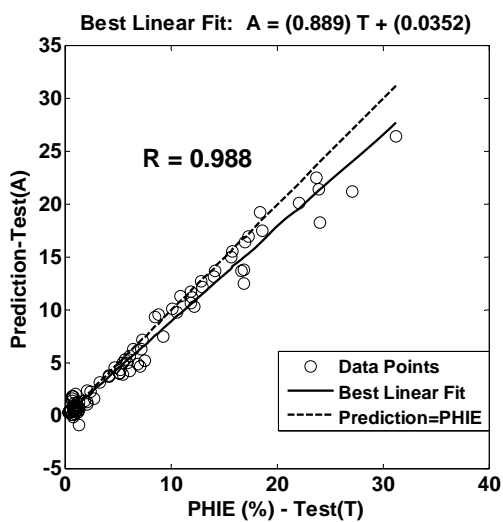
تعداد کل پارامترهای شبکه	ضریب همبستگی (الگوهای آزمون)	MSE (الگوهای آزمون)	ضریب همبستگی (الگوهای آموزش)	MSE (الگوهای آموزش)	ساختار شبکه	ردیف
۸۷	۰/۹۸۸	۲/۲۴۳	۰/۹۹۲	۱/۱۳۶	۴-۲-۱۹-۱	۱
۲۶	۰/۹۸۷	۲/۲۵۸	۰/۹۹۱	۱/۳۲	۴-۳-۲-۱	۲
۵۵	۰/۹۸۷	۲/۲۸۸	۰/۹۹۳	۱/۰۶۶	۴-۲-۱۱-۱	۳
۳۱	۰/۹۸۷	۲/۳۳۴	۰/۹۹۲	۱/۱۰۴	۴-۳-۳-۱	۴
۳۱	۰/۹۸۶	۲/۳۷	۰/۹۹۱	۱/۲۹۷	۴-۲-۵-۱	۵
۷۹	۰/۹۸۴	۲/۳۷۴	۰/۹۹۳	۱/۰۵۵	۴-۲-۱۷-۱	۶
۴۱	۰/۹۸۶	۲/۳۷۵	۰/۹۹۳	۰/۹۷۵	۴-۳-۵-۱	۷

در نمودارهای همبستگی، معادله بهترین خط برازش شده بین نقاط و همچنین مقدار ضریب همبستگی (R) با نرم‌افزار مطلب محاسبه و نشان داده شده است.

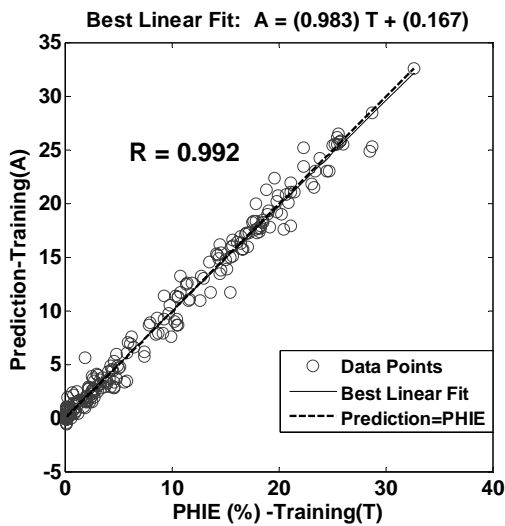
شکل ۶ پیشرفت روند آموزش و نمودار همبستگی (Cross plot) بین تخلخل و برآورد بهترین شبکه منفرد حاصل (شبکه شماره ۱ جدول ۱) را برای الگوهای آموزش و آزمون نشان می‌دهد.



(الف)



(ج)



(ب)

شکل ۶. الف) نمایش پیشرفت روند آموزش و کاهش MSE برای الگوهای آموزش در آموزش شبکه شماره ۱ جدول ۱. ب) نمودار همبستگی بین تخلخل و برآورد شبکه شماره ۱ جدول ۱ برای الگوهای آموزش. ج) نمودار همبستگی بین تخلخل و برآورد شبکه شماره ۱ جدول ۱ برای الگوهای آزمون.

۱) بچه‌های تولید شده با استفاده از قانون تقاطع را تعیین می‌کند از ۰/۳ تا ۰/۸ تغییر داده شد. به گزینه‌های قابل انتخاب مربوط به الگوریتم ژنتیک مقادیر و موارد متفاوتی نسبت داده شد و فرایند آزمون و خطا با جمعیت‌های اولیه تصادفی گوناگون صورت گرفت. در نهایت در هر یک از شش حالت منتخب، بهترین نتیجه که کمترین میانگین مربعات خطای برآورد تخلخل الگوهای آزمون را داشت ثبت شد. نتایج حاصل به‌طور مشروح در زیر آمده است.

نتایج مربوط به تنها ترکیب هفت شبکه‌ای که همه هفت شبکه جدول ۱ را شامل می‌شود در جدول ۲ آورده شده است. ستون سوم از سمت راست این جدول مربوط به ترکیب خطی بهینه نامقید بدون جمله ثابت است که در بین پنج روش تحلیلی، بهترین نتیجه، یعنی کمترین میانگین مربعات خطای برآورد تخلخل الگوهای آزمون را داشته است. در این جدول و جدول‌های مشابه که در ادامه می‌آید، درصد کاهش MSE برآورد تخلخل الگوهای آموزش از تقسیم MSE برآورد ترکیب بر MSE برآورد بهترین شبکه منفرد برای الگوهای آموزش و تفریق مقدار حاصل از ۱ به دست آمده است. درصد کاهش MSE برآورد تخلخل الگوهای آزمون نیز به‌طور مشابه محاسبه شده است.

۱۲۰ ترکیب آنسامبلی خطی ممکن این ۷ شبکه (یعنی ۲۱ ترکیب دوشبکه‌ای، ۳۵ ترکیب سه‌شبکه‌ای، ۳۵ ترکیب چهارشبکه‌ای، ۲۱ ترکیب پنج‌شبکه‌ای، ۷ ترکیب شش‌شبکه‌ای و ۱ ترکیب هفت‌شبکه‌ای) با هر یک از ۵ روش تحلیلی میانگین‌گیری ساده، ترکیب خطی بهینه نامقید با جمله ثابت، ترکیب خطی بهینه مقید با جمله ثابت، ترکیب خطی بهینه نامقید بدون جمله ثابت و ترکیب خطی بهینه مقید بدون جمله ثابت به دست آمد. سپس بهترین ترکیب حاصل از ۵ روش تحلیلی از بین هر یک از دسته ترکیب‌های دو، سه، چهار، پنج، شش و هفت شبکه‌ای که کمترین میانگین مربعات خطای برآورد تخلخل الگوهای آزمون را در دسته داشت، انتخاب شد. برای این شش ترکیب منتخب، علاوه بر روش‌های تحلیلی، ضرایب ترکیب خطی بهینه با استفاده از روش الگوریتم ژنتیک تعیین و ترکیب آنسامبلی خطی ساخته شد. در بهینه‌یابی با الگوریتم ژنتیک، تعداد کروموزم‌های جمعیت‌ها ۵۰ کروموزوم انتخاب شد که از این تعداد، ۵ کروموزوم نخبه در جمعیت بعدی حفظ و مابقی با استفاده از تقاطع و جهش تولید شد. تعداد دفعات تولید نسل ۱۵۰ یا ۲۰۰ بار انتخاب شده است. مقادیر کمیت نسبت تقاطع (Crossover fraction) که درصد یا مقدار نسبی (نسبتی از

جدول ۲. نتایج ترکیب آنسامبلی خطی هفت شبکه جدول ۱.

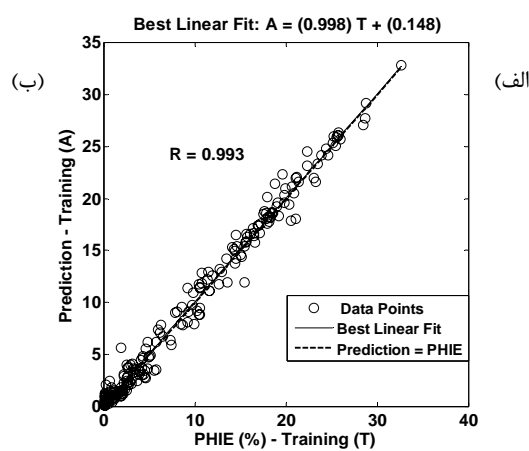
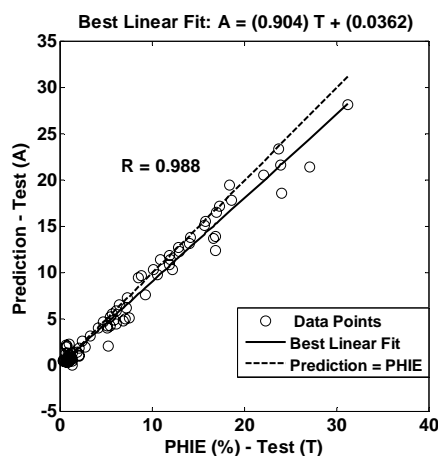
GA - OLC	Hashem - OLC	میانگین‌گیری ساده	روش ترکیب خطی نتایج
۷ و ۶، ۵، ۴، ۳، ۲، ۱	۷ و ۶، ۵، ۴، ۳، ۲، ۱	۷ و ۶، ۵، ۴، ۳، ۲، ۱	ترکیب آنسامبلی
۰/۹۷۵	۰/۸۸۴	۰/۹۸۷	MSE (الگوهای آموزش)
۱۴/۲٪	۲۲/۲٪	۱۳/۱٪	کاهش در الگوهای آموزش.
۱/۹۹۳	۲/۱۹۴	۲/۲۲۳	MSE (الگوهای آزمون)
۱۱/۱٪	۲/۲٪	۰/۹٪	کاهش در الگوهای آزمون.

استفاده از الگوریتم ژنتیک را برای الگوهای آموزش و آزمون نشان می‌دهد.

از بین هفت ترکیب شش شبکه‌ای ممکن به ۵ روش تحلیلی (میانگین‌گیری ساده و ۴ روش ترکیب خطی بهینه با روش هاشم)، یعنی مجموعاً ۳۵ ترکیب، بهترین نتیجه ترکیب شبکه‌های ۱، ۲، ۳، ۴، ۵ و ۶ به روش نامقید بدون جمله ثابت بوده است که نسبت به بهترین شبکه منفرد (شبکه ۱ جدول ۱) بیشترین کاهش را در میانگین مربعات خطای برآورد تخلخل الگوهای آزمون ایجاد کرده است. به منظور بهینه‌سازی ترکیب خطی این شش شبکه، روش الگوریتم ژنتیک نیز به کار گرفته شد. جدول ۳ نتایج میانگین‌گیری ساده، ترکیب خطی بهینه نامقید بدون جمله ثابت و ترکیب خطی بهینه با استفاده از الگوریتم ژنتیک برای ترکیب آنسامبلی متشکل از این ۶ شبکه را نشان می‌دهد. این مورد نیز نتایج مشابه با حالت قبلی دارد.

با توجه به مقادیر موجود در جدول ۲ مشاهده می‌شود که ترکیب آنسامبلی شبکه‌های منفرد توانسته است خطا را کاهش و نتایج را بهبود دهد. بهترین ترکیب آنسامبلی خطی هفت شبکه‌ای حاصل، مربوط به ترکیب با استفاده از الگوریتم ژنتیک است که از MSE برآورد تخلخل الگوهای آزمون بیشترین مقدار یعنی ۱۱/۱٪ کاسته است. همچنین اگر چه روش‌های تحلیلی، MSE برآورد تخلخل الگوهای آموزش و آزمون را کاهش داده‌اند اما MSE برآورد تخلخل الگوهای آزمون را به مراتب کمتر از الگوهای آموزش کاهش داده‌اند؛ درحالی‌که ترکیب آنسامبلی خطی بهینه با استفاده از الگوریتم ژنتیک MSE برآورد تخلخل الگوهای آزمون را نیز به خوبی کاهش داده است.

شکل ۷ نمودار همبستگی بین تخلخل و برآورد ترکیب آنسامبلی هفت شبکه‌ای به روش خطی بهینه با



شکل ۷. نمودار همبستگی بین تخلخل و برآورد ترکیب آنسامبلی هفت شبکه‌ای به روش خطی بهینه با استفاده از الگوریتم ژنتیک برای الف) الگوهای آموزش و ب) الگوهای آزمون.

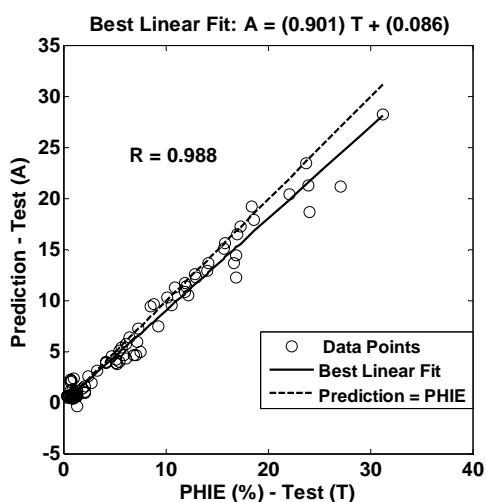
جدول ۳. نتایج ترکیب آنسامبلی خطی شش شبکه ۱، ۲، ۳، ۴، ۵ و ۶.

GA - OLC	Hashem - OLC	میانگین‌گیری ساده	روش ترکیب خطی
			نتایج
۱، ۲، ۳، ۴، ۵ و ۶	۱، ۲، ۳، ۴، ۵ و ۶	۱، ۲، ۳، ۴، ۵ و ۶	ترکیب آنسامبلی
۰/۹۷۲	۰/۹۰۷	۱/۰۱۵	MSE (الگوهای آموزش)
۱۴/۴٪	۲۰/۲٪	۱۰/۷٪	کاهش در MSE الگوهای آموزش.
۱/۹۶۳	۲/۱۸۴	۲/۲۰۸	MSE (الگوهای آزمون)
۱۲/۵٪	۲/۶٪	۱/۶٪	کاهش در MSE الگوهای آزمون.

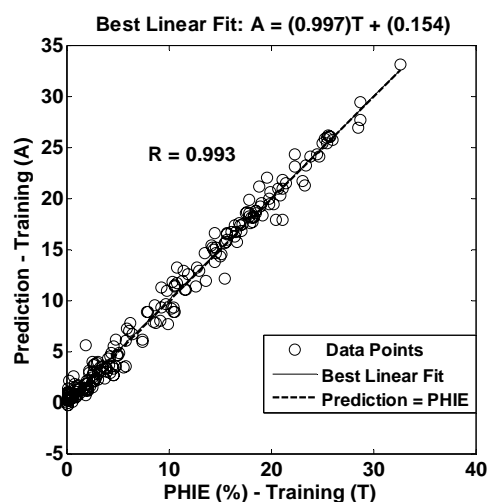
مربوط به ترکیب شبکه‌های ۱، ۲، ۳، ۴ و ۵ به روش نامقید بدون جمله ثابت است که کمترین میانگین مربعات خطای برآورد تخلخل الگوهای آزمون را دارد. ترکیب خطی بهینه این پنج شبکه با الگوریتم ژنتیک نیز به دست آمد. نتایج در جدول ۴ آورده شده است. نتایج این جدول نشان می‌دهد که بهترین ترکیب پنج شبکه‌ای حاصل با استفاده از الگوریتم ژنتیک به دست آمده است، زیرا بیشترین کاهش به میزان ۹/۲٪ را در میانگین مربعات خطای برآورد تخلخل الگوهای آزمون ایجاد کرده است.

نمودار همبستگی بین تخلخل و برآورد ترکیب آنسامبلی شش شبکه‌ای شبکه‌های ۱، ۲، ۳، ۴، ۵ و ۶ به روش خطی بهینه با استفاده از الگوریتم ژنتیک برای الگوهای آموزش و آزمون در شکل ۸ نشان داده شده است.

ترکیبات پنج شبکه‌ای از بین هفت شبکه جدول ۱ را می‌توان به ۲۱ حالت انتخاب کرد. ترکیبات آنسامبلی خطی این ۲۱ حالت به ۵ روش تحلیلی، یعنی مجموعاً ۱۰۵ ترکیب، به دست آمد که از بین آنها بهترین ترکیب



(ب)



(الف)

شکل ۸. نمودار همبستگی بین تخلخل و برآورد ترکیب آنسامبلی شش شبکه‌ای شبکه‌های ۱، ۲، ۳، ۴، ۵ و ۶ به روش خطی بهینه با استفاده از الگوریتم ژنتیک برای الف) الگوهای آموزش و ب) الگوهای آزمون.

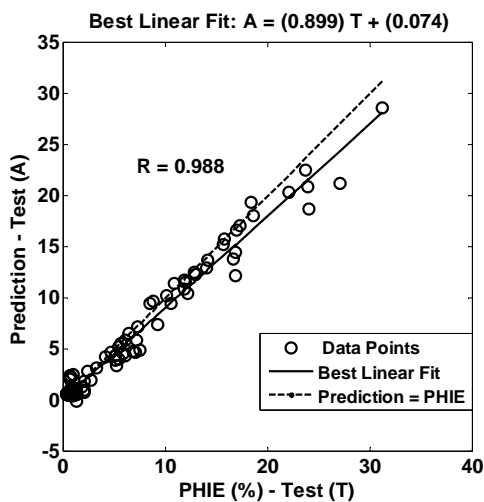
جدول ۴. نتایج ترکیب آنسامبلی خطی پنج شبکه ۱، ۲، ۳، ۴ و ۵.

GA - OLC	Hashem - OLC	میانگین گیری ساده	روش ترکیب خطی نتایج
۵ و ۴، ۳، ۲، ۱	۵ و ۴، ۳، ۲، ۱	۵ و ۴، ۳، ۲، ۱	ترکیب آنسامبلی
۱/۰۰۷	۰/۹۵۱	۱/۰۴۶	MSE (الگوهای آموزش)
۱۱/۴٪	۱۶/۳٪	۷/۹٪	کاهش در MSE الگوهای آموزش.
۲/۰۳۶	۲/۱۷۴	۲/۲۰۹	MSE (الگوهای آزمون)
۹/۲٪	۳/۱٪	۱/۵٪	کاهش در MSE الگوهای آزمون.

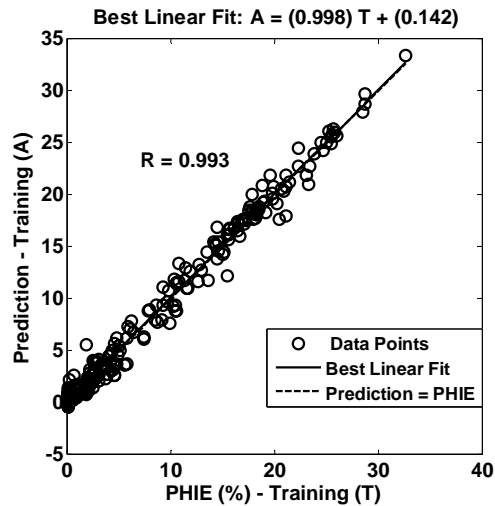
آنسامبلی با روش‌های میانگین‌گیری ساده، ترکیب خطی بهینه نامقید بدون جمله ثابت و ترکیب خطی بهینه با الگوریتم ژنتیک را برای این چهار شبکه نشان می‌دهد. در این مورد نیز اگرچه ترکیب خطی بهینه نامقید بدون جمله ثابت بیشترین کاهش را در MSE برآورد تخلخل الگوهای آموزش ایجاد کرده است ولی ترکیب خطی بهینه به دست آمده با استفاده از الگوریتم ژنتیک بیشترین کاهش را در MSE برآورد تخلخل الگوهای آزمون توان تعمیم در مقایسه با بهترین شبکه عصبی منفرد حاصل به‌وجود آورده است.

شکل ۹ نمودار همبستگی بین تخلخل و برآورد ترکیب آنسامبلی پنج‌شبکه‌ای شبکه‌های ۱، ۲، ۳، ۴ و ۵ به روش خطی بهینه با استفاده از الگوریتم ژنتیک را برای الگوهای آموزش و آزمون نشان می‌دهد.

۳۵ انتخاب چهارشبکه‌ای ممکن از بین هفت شبکه صورت گرفت و با استفاده از ۵ روش تحلیلی مجموعاً ۱۷۵ ترکیب به دست آمد که بهترین نتیجه که کمترین میانگین مربعات خطای برآورد تخلخل الگوهای آزمون را دارد مربوط به ترکیب خطی بهینه نامقید بدون جمله ثابت شبکه‌های ۱، ۲، ۳ و ۵ است. جدول ۵ نتایج ترکیب



(ب)



(الف)

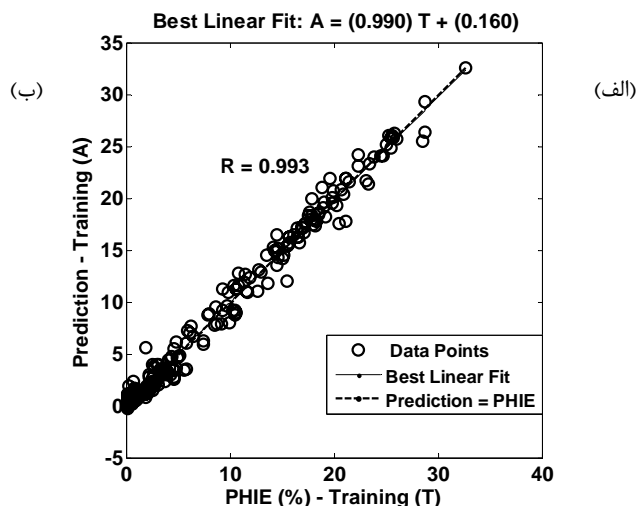
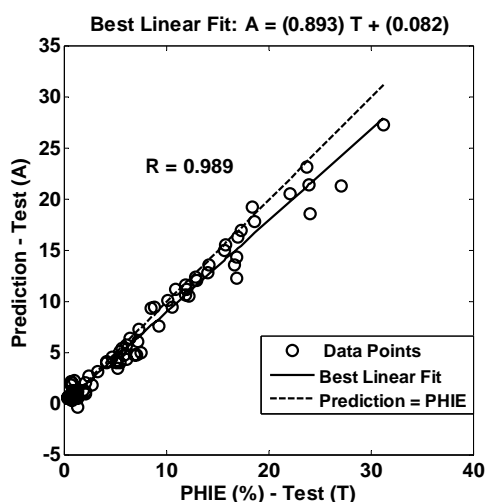
شکل ۹. نمودار همبستگی بین تخلخل و برآورد ترکیب آنسامبلی پنج‌شبکه‌ای شبکه‌های ۱، ۲، ۳، ۴ و ۵ به روش خطی بهینه با استفاده از الگوریتم ژنتیک برای (الف) الگوهای آموزش و (ب) الگوهای آزمون.

جدول ۵. نتایج ترکیب آنسامبلی خطی چهار شبکه ۱، ۲، ۳ و ۵.

GA - OLC	Hashem - OLC	میانگین‌گیری ساده	روش ترکیب خطی نتایج
۵ و ۳، ۲، ۱	۵ و ۳، ۲، ۱	۵ و ۳، ۲، ۱	ترکیب آنسامبلی
۰/۹۹۹	۰/۹۶۴	۱/۰۶۸	MSE (الگوهای آموزش)
۱۲/۱٪	۱۵/۱٪	۶٪	کاهش در الگوهای آموزش.
۲/۰۳۵	۲/۱۴۸	۲/۱۹۵	MSE (الگوهای آزمون)
۹/۳٪	۴/۲٪	۲/۱٪	کاهش در الگوهای آزمون.

جمله ثابت و ترکیب خطی بهینه با الگوریتم ژنتیک در جدول ۶ آورده شده است. در مورد ترکیب شبکه‌های ۱، ۲ و ۵ با روش میانگین‌گیری ساده، در مورد الگوهای آموزش بهبودی حاصل نشده که این با علامت (---) نشان داده شده است. مقادیر مندرج در جدول ۶ نشان می‌دهد که در ترکیب با روش‌های میانگین‌گیری ساده و GA-OLC، MSE برآورد تخلخل الگوهای آزمون بیشتر از الگوهای آموزش کاهش یافته است.

نمودار همبستگی بین تخلخل و برآورد ترکیب آنسامبلی چهارشبکه‌ای شبکه‌های ۱، ۲، ۳ و ۵ به روش خطی بهینه با استفاده از الگوریتم ژنتیک برای الگوهای آموزش و آزمون در شکل ۱۰ نشان داده شده است. در مورد انتخاب سه شبکه از بین هفت شبکه نیز ۳۵ حالت وجود دارد که بهترین نتیجه روش‌های تحلیلی، مربوط به ترکیب خطی بهینه نامقید بدون جمله ثابت سه شبکه ۱، ۲ و ۵ است. نتایج ترکیب خطی این سه شبکه با روش میانگین‌گیری ساده، ترکیب خطی بهینه نامقید بدون



شکل ۱۰. نمودار همبستگی بین تخلخل و برآورد ترکیب آنسامبلی چهارشبکه‌ای شبکه‌های ۱، ۲، ۳ و ۵ به روش خطی بهینه با استفاده از الگوریتم ژنتیک برای الف) الگوهای آموزش و ب) الگوهای آزمون.

جدول ۶. ترکیب آنسامبلی سه شبکه ۱، ۲ و ۵.

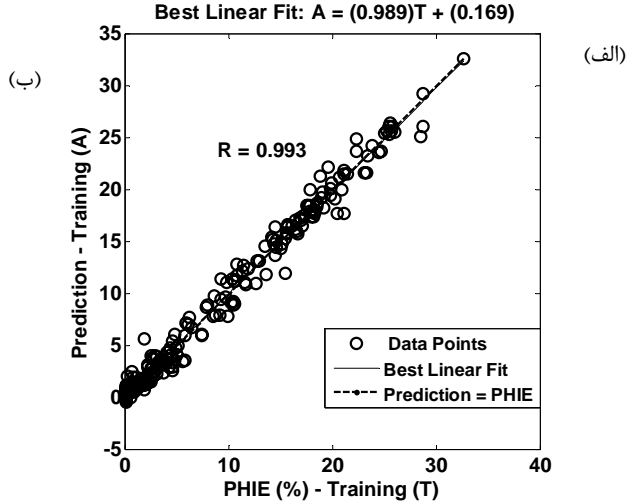
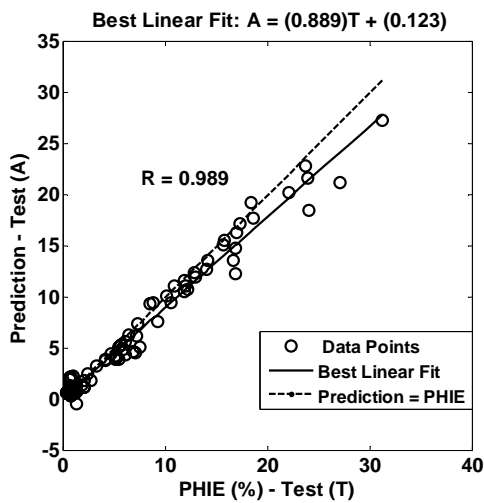
GA - OLC	Hashem - OLC	میانگین‌گیری ساده	روش ترکیب خطی نتایج
۵ و ۲، ۱	۵ و ۲، ۱	۵ و ۲، ۱	ترکیب آنسامبلی
۱/۰۶۹	۱/۰۶۴	۱/۱۴۳	MSE (الگوهای آموزش)
۵/۹٪	۶/۳٪	---	کاهش در MSE الگوهای آموزش.
۲/۰۴۹	۲/۱۳۳	۲/۱۹۸	MSE (الگوهای آزمون)
۸/۶٪	۴/۹٪	۲٪	کاهش در MSE الگوهای آزمون.

برآورد تخلخل الگوهای آزمون و بهترین نتیجه را داشت. ترکیب خطی بهینه شبکه‌های ۱ و ۲ با استفاده از الگوریتم ژنتیک نیز به دست آمد. جدول ۷ نتایج این مورد را نشان می‌دهد.

نمودار همبستگی بین تخلخل و برآورد ترکیب آنسامبلی دوشبکه‌ای شبکه‌های ۱ و ۲ به روش خطی بهینه با استفاده از الگوریتم ژنتیک برای الگوهای آموزش و آزمون در شکل ۱۲ نشان داده شده است.

شکل ۱۱ نمودار همبستگی بین تخلخل و برآورد ترکیب آنسامبلی سه‌شبکه‌ای شبکه‌های ۱، ۲ و ۵ به روش خطی بهینه با استفاده از الگوریتم ژنتیک را برای الگوهای آموزش و آزمون نشان می‌دهد.

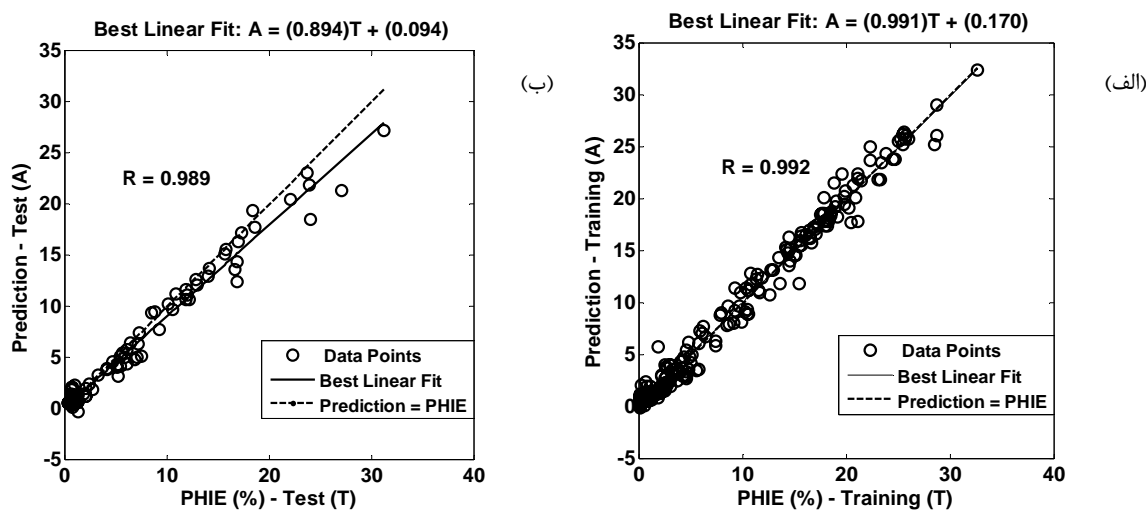
۵ روش تحلیلی در مورد ۲۱ انتخاب دوشبکه‌ای ممکن از بین هفت شبکه جدول ۱ به کار رفت که از بین ۱۰۵ ترکیب حاصل، ترکیب خطی بهینه نامقید بدون جمله ثابت دو شبکه ۱ و ۲ کمترین میانگین مربعات خطای



شکل ۱۱. نمودار همبستگی بین تخلخل و برآورد ترکیب آنسامبلی سه‌شبکه‌ای شبکه‌های ۱، ۲ و ۵ به روش خطی بهینه با استفاده از الگوریتم ژنتیک برای الف) الگوهای آموزش و ب) الگوهای آزمون.

جدول ۷. ترکیب آنسامبلی دو شبکه ۱ و ۲.

GA - OLC	Hashem - OLC	میانگین گیری ساده	روش ترکیب خطی نتایج
۲ و ۱	۲ و ۱	۲ و ۱	ترکیب آنسامبلی
۱/۰۹۷	۱/۰۸۶	۱/۱۰۳	MSE (الگوهای آموزش)
۳/۴٪	۴/۴٪	۲/۹٪	کاهش در MSE الگوهای آموزش.
۲/۰۱۳	۲/۱۵۰	۲/۱۵۶	MSE (الگوهای آزمون)
۱۰/۳٪	۴/۱٪	۳/۹٪	کاهش در MSE الگوهای آزمون.



شکل ۱۲. نمودار همبستگی بین تخلخل و برآورد ترکیب آنسامبلی دوشبکه‌ای شبکه‌های ۱ و ۲ به روش خطی بهینه با استفاده از الگوریتم ژنتیک برای الف) الگوهای آموزش و ب) الگوهای آزمون.

خطی بهینه با روش الگوریتم ژنتیک نتایج بهتری داشته است؛ زیرا کاهش بیشتری در MSE برآورد تخلخل الگوهای آزمون ایجاد کرده است. این درحالی است که روش تحلیلی هاشم بیشترین کاهش را در MSE برآورد تخلخل الگوهای آموزش به وجود آورده است.

به‌طورکلی میزان توانایی ترکیب آنسامبلی در بهبود نتایج شبکه‌های عصبی مصنوعی منفرد و افزایش دقت و کاهش خطا در حل مسئله‌ها تابع عواملی است. اولین عامل مؤثر در این زمینه، نوع و ماهیت مسئله است، به این مفهوم که چه نوع کمیت‌هایی ورودی شبکه‌ها هستند و کمیت‌های خروجی چه ماهیتی دارند؟ شبکه‌های منفردی که از آنها در حکم اجزای آنسامبل استفاده می‌شود نیز در کارایی ترکیب آنسامبلی تأثیر زیادی دارند. از این منظر، مجموعه الگوهای آموزشی، ساختار، پارامترهای اولیه و روش آموزش شبکه‌های جزء اهمیت پیدا می‌کند. عامل سومی که می‌توانیم در اینجا به آن اشاره کنیم روش ترکیبی است که از آن برای ساخت ترکیبات بهره می‌گیریم. چنان‌که دیدیم ترکیب با روش‌های میانگین‌گیری ساده، روش ترکیب خطی بهینه هاشم و

۷ نتیجه‌گیری

نتایج حاصل از به کارگیری ترکیب آنسامبلی در تحقیق موردی صورت گرفته در این تحقیق نشان می‌دهد که ترکیب آنسامبلی توانسته است نتایج شبکه‌های عصبی مصنوعی منفرد را به خوبی بهبود دهد. از بین همه ترکیبات آنسامبلی خطی ساخته شده، ترکیب خطی بهینه شش شبکه ۱، ۲، ۳، ۴، ۵ و ۶ جدول ۱ با استفاده از الگوریتم ژنتیک بهترین نتیجه را داشته است زیرا MSE برآورد تخلخل الگوهای آزمون را بیشترین مقدار یعنی ۱۲/۵٪ نسبت به بهترین شبکه منفرد حاصل (شبکه شماره ۱ جدول ۱) کاهش داده است. این ترکیب از MSE برآورد تخلخل الگوهای آموزش ۱۴/۴٪ نسبت به بهترین شبکه منفرد کاسته است. از بین ترکیب‌های حاصل از روش‌های تحلیلی، بهترین نتیجه مربوط به ترکیب نامقید بدون جمله ثابت شبکه‌های ۱، ۲ و ۵ جدول ۱ است که MSE برآورد تخلخل الگوهای آزمون را در مقایسه با سایر ترکیب‌های ناشی از روش‌های تحلیلی بیشترین مقدار یعنی ۴/۹٪ نسبت به بهترین شبکه منفرد کاهش داده است. به‌طورکلی در همه حالت‌هایی که مورد بررسی قرار گرفته، ترکیب

- Department of Petroleum Engineering and Applied Geophysics, Norwegian University of Science and Technology.
- Bhatt A. and Helle H.B., 2002, Committee neural networks for porosity and permeability prediction from well logs. *Geophysical Prospecting*, **50**, 645-660.
- Chen Ch.H. and Lin Z.Sh., 2006, A committee machine with empirical formulas for permeability prediction. *Computer & Geosciences*, **32**, 485-496.
- Fung Ch.Ch., Wong K.W. and Eren H., 1997, Modular artificial neural network for prediction of petrophysical properties from well log data. *IEEE transaction on instrumentation and measurement*, **46**, 1295-1299.
- Hashem S., 1993, Optimal linear combination of neural networks. PhD thesis, School of Industrial Engineering, Purdue University.
- Hashem S., 1997, Optimal linear combination of neural networks. *Neural Networks*, **10**, 599-614.
- Hashem S., Shmeiser B. and Yih Y., 1994, Optimal linear combination of neural networks: An overview. School of Industrial Engineering, Purdue University.
- Haykin S., 1999, *Neural networks: A comprehensive foundation*. Prentice-Hall, Inc.
- Hearst J.R., Nelson P.H. and Paillet F.L., 2000, *Well logging for physical properties, A hand book for geophysicists, geologists and engineers*. John Wiley and Sons Ltd.
- Helle H.B., Bhatt A. and Ursin B., 2001, Porosity and permeability prediction from wireline logs using artificial neural networks: A North Sea case study. *Geophysical Prospecting*, **49**, 431-444.
- Huang Z., Shimeld J., Williamson M. and Katsube J., 1996, Permeability prediction with artificial neural neural network modeling in the Venture gas field offshore eastern Canada. *Geophysics*, **61**, 422-436.
- Huang Z. and Williamson M., 1997, Determination of porosity and permeability in reservoir intervals by artificial neural network modelling offshore eastern Canada. *Petroleum Geoscience*, **3**, 245-258.
- Naftaly U., Intrator N. and Horn D., 1997, Optimal ensemble averaging of neural networks. *Computation of Neural Systems*, **8**, 283-296.
- Parmanto B., Munro P.W. and Doyle H.R., 1996, Reducing variance of committee predictions
- ترکیب خطی با استفاده از الگوریتم ژنتیک نتایج متفاوتی دارند. به‌هرحال توانایی (یا ناتوانی) ترکیب آنسامبلی در بهبود نتایج شبکه‌های عصبی مصنوعی منفرد و همچنین مقدار بهبود صورت گرفته، از یک مسئله به مسئله دیگر متفاوت است و بنابراین در این مورد نمی‌توان یک اظهارنظر یا نتیجه‌گیری کلی ابراز داشت.
- ### تشکر و قدردانی
- از آقای دکتر پویان، معاون محترم پژوهشی دانشگاه صنعتی شاهرود به‌خاطر راهنمایی‌های ارزنده‌شان و همچنین از آقای مهندس ماشاء... رحیمی، کارشناس ارشد مدیریت اکتشاف شرکت ملی نفت به‌خاطر در اختیار قرار دادن داده‌های مورد نیاز این مقاله و راهنمایی‌های با ارزش به نگارندگان، تشکر و قدردانی می‌شود.
- ### منابع
- افشارحرب، ع.، ۱۳۸۱، زمین‌شناسی نفت، انتشارات دانشگاه پیام نور.
- حسینی پاک، ع.ا. و شرف‌الدین، م.، ۱۳۸۰، تحلیل داده‌های اکتشافی، انتشارات دانشگاه تهران.
- خسروتهرانی، خ.، ۱۳۸۳، زمین‌شناسی ایران، انتشارات دانشگاه پیام نور، چاپ پنجم.
- رحیمی، م.، ۱۳۸۱، تعبیر و تفسیر لرزه‌ای ناحیه پارس جنوبی در افق‌های آسماری، جهرم، سروک، داریان، فهلیان، سورمه، کنگان، نار و فراقون، مدیریت اکتشاف شرکت ملی نفت ایران.
- زارعی، س.، موحد، ب.، باقری، ع.م. و مردانی، ع.، ۱۳۸۶، ارزیابی تراوایی سازند کنگان با استفاده از داده‌های نگار CMR و مغزه در میدان پارس جنوبی، بیست و ششمین گردهمایی علوم زمین‌شناسی.
- Bhatt A., 2002, Reservoir properties from well logs using neural networks. PhD thesis,

with resampling techniques. *Connection Science*, **8**, 405-425.

Sharkey A., Sharkey N., Gerecke U. and Chandroth G.O., 2000, The "test and select" approach to ensemble combination. *In* Multiple classifier systems (Eds J.Kittler and F.Roli), 30-44, Lecture notes in computer science, Springer-Verlag, Inc.