

توسعه مدل شبکه عصبی بر مبنای توابع آموزش گرادیان مزدوج و پس انتشار ارتجاعی برای پیش بینی ضریب انتشار طولی رودخانه‌ها

بهزاد قیاسی^۱، روح‌اله نوری^{۲*}، عبدالرضا کرباسی^۲، امین سارنگ^۲

چکیده

گام اساسی در مدل‌سازی کیفی محیط‌های آبی یک بعدی مانند رودخانه‌ها، تعیین ضریب انتشار طولی (LDC) برای معادله‌ی انتقال-پخش آلاینده‌ها است. در این مقاله برای پیش‌بینی LDC، مدل شبکه‌ی عصبی مصنوعی (ANN) بر مبنای الگوریتم‌های آموزشی با رویکرد عددی و همچنین رویکرد اکتشافی توسعه داده شده است. برای این منظور توابع آموزشی گرادیان مزدوج شامل توابع فلچر-ریوس، پولاک-ریبره، پاول-بیل و گرادیان مزدوج مقیاس‌دار از دسته الگوریتم‌های عددی و همچنین تابع پس‌انتشار ارتجاعی از دسته الگوریتم‌های اکتشافی برای بهینه‌سازی پارامترهای مدل ANN استفاده شدند. در مرحله‌ی بعد با استفاده از آماره‌های بررسی شده برای ارزیابی نتایج، بهترین مدل با ساختار شامل هر یک از توابع نامبرده انتخاب شدند و در ادامه از بین مدل‌های منتخب، مدلی که بهترین عملکرد را داشت، یعنی مدل با تابع آموزش پس‌انتشار ارتجاعی، با توجه به آماره‌ی نسبت تفاوت توسعه یافته (DDR)، به عنوان نتیجه نهایی این مقاله برگزیده شد. در پایان نیز برای ارزیابی بهتر نتایج تحقیق، رویکردی مقایسه‌ای بین نتیجه بهترین مدل توسعه داده شده با دیگر مطالعات انجام گرفته به وسیله مدل‌های هوشمند انجام شد که یافته‌ها حاکی از عملکرد برتر مدل پس‌انتشار ارتجاعی بود.

واژه‌های کلیدی: ضریب پخش طولی، آلودگی آب، مدل‌های هوشمند، الگوریتم آموزش

^۱ دانشجوی دکتری مهندسی محیط‌زیست-منابع آب، دانشکده فنی، دانشگاه تهران

^۲ استادیار مهندسی محیط‌زیست، دانشکده فنی، دانشگاه تهران

تلفن: ۰۹۱۲۸۳۵۲۴۸۸ Email: noor@ut.ac.ir

^۳ دانشیار مهندسی محیط‌زیست، دانشکده فنی، دانشگاه تهران

مقدمه

وقتی یک آلاینده به جریانی وارد می‌شود، فرایندهایی را که رخ می‌دهد، می‌توان به سه بخش تقسیم کرد. در بخش اول اندازه‌ی حرکت و شناوری جریان ورودی، نرخ رقیق‌سازی را مشخص می‌کند. هنگامی که آلاینده رقیق شد، اثر نیروهای اندازه حرکت و شناوری جریان نیز کاهش می‌یابد و به مرحله‌ی دوم هدایت می‌گردد که در آن آلاینده به واسطه‌ی آشفتگی در مقطع جریان پخش می‌شود. در نهایت وقتی آلاینده به طور کامل در مقطع جریان پخش شد، فرایند پخش طولی آلاینده در رودخانه حاکم می‌گردد. در واقع این سه قسمت را می‌توان به سه بخش دسته‌بندی کرد: (۱) پخش در نزدیکی محل رها شدن ماده بر اثر نیروی اندازه‌ی حرکت و شناوری اولیه، (۲) پخش جانبی بر اثر پخش انتقالی، و (۳) پخش برشی طولی جریان. دو مرحله‌ی اول در نزدیکی محل ورود آلاینده به جریان رخ می‌دهند و عملاً فرایند اصلی رقیق‌سازی و عامل تعیین کننده‌ی مقدار آن، ضریب انتشار طولی^۱ (LDC) است (رادرفورد، ۱۹۹۴). اهمیت LDC در معادله انتقال - پخش نیز به وضوح قابل درک است، زیرا مقدار غلظت در زمان و مکان مطلوب وابسته به مقدار LDC است. بنابراین تعیین LDC در رودها، می‌تواند کمک شایانی به مدل‌سازی رفتار آلاینده‌ها جهت اتخاذ تدابیر لازم برای مقابله با آنها نماید. تخمین دقیق LDC در رودها نیازمند طراحی مطالعات آزمایشگاهی و تجربی فراوانی است که این امر با هزینه‌های مالی و زمانی زیادی همراه است. بنابراین در راستای کمی کردن این پارامتر به عنوان یک راهکار جایگزین، محققان زیادی تا به حال در پی ارائه روابط تجربی بر پایه‌ی مشخصات هیدرولیکی و هندسی جریان، و همچنین استفاده از مدل‌های هوشمند بوده‌اند. فیشر (۱۹۷۵) رابطه‌ای تجربی بر مبنای مشخصات هیدرولیکی و هندسی رودخانه ارائه داد که این رابطه اگرچه برای تخمین اولیه LDC مناسب است، ولی خطای زیاد در تخمین پارامتر مزبور از معایب آن است. در تحقیقی سئو و چونگ (۱۹۹۸) تلاش کردند تا با استفاده از روش رگرسیونی یک مرحله‌ای هابر (۱۹۸۱) مدلی مناسب برای

تعیین LDC در رودخانه‌های طبیعی ارائه دهند. کاشفی‌پور (۲۰۰۷) با استفاده از شبکه عصبی مصنوعی (ANN)^۲ و به‌کارگیری پارامترهای بدون بعد شده‌ی هیدرولیکی و هندسی جریان، اقدام به پیش‌بینی LDC در مجاری روباز کرد. وی در نهایت با مقایسه نتایج خود با مطالعه سئو و چونگ (۱۹۹۸) و کاشفی‌پور و فالدکنر (۲۰۰۲) به این نتیجه رسید که مدل پیشنهادی وی از دقت بالاتری (حدود ۱۴ درصد) برخوردار است. ریاحی مدوار و ایوب‌زاده (۲۰۰۸) با بهره‌گیری از سامانه استنتاج فازی-عصبی تطبیقی^۳ (ANFIS)، روشی نوینی را برای تخمین LDC ارائه نمودند که در مقایسه با روابط تجربی، عملکرد به مراتب بهتری از خود نشان می‌داد. گیوه‌چی و همکاران (۲۰۰۹) با استفاده از سرعت متوسط عمقی، LDC را تخمین زدند که نتایج حاکی از دقت خوب این روش بود. نوری و همکاران (۲۰۰۹) در تحقیقی استفاده از مدل‌های ماشین بردار پشتیبان^۴ (SVM) و ANFIS را برای تخمین با دقت بالای LDC پیشنهاد نمودند. آداریش (۲۰۱۰) نیز با استفاده از مدل‌های هوشمند SVM و برنامه‌نویسی ژنتیک^۵، به دقت مناسبی از تخمین LDC رسید که برتری قابل ملاحظه‌ای را در مقابل مدل‌های تجربی داشت. اکبرزاده و همکاران (۲۰۱۰) به بررسی عدم قطعیت مدل‌های هوشمند در تخمین LDC با استفاده از روش مونت-کارلو پرداختند. اعتماد شهیدی و تقی‌پور (۲۰۱۲) با استفاده از مدل درختی M5، مدلی جدید برای تخمین LDC جریان‌های طبیعی ارائه کردند. پارسایی و حقی‌آبی (۲۰۱۳) با ارزیابی روابط تجربی و مدل‌های هوش مصنوعی ارائه شده برای تخمین LDC در دو رود Narew و Severn، برتری مدل‌های هوشمند را نتیجه گرفتند. توتمز (۲۰۱۳) از روش رگرسیونی کریجینگ^۶ بهره برد و دو مدل از این روش ارائه کرد که دقت بهتری نسبت به مدل‌های تحقیقات نوری (۲۰۰۹)، کاشفی‌پور (۲۰۰۲) و سئو و چونگ (۱۹۹۸) را نشان می‌داد. ژنگ (۲۰۱۴) با ارائه رابطه‌ای تجربی بر مبنای مشخصات هیدرولیکی و هندسی جریان، به دقت مناسبی دست یافت. نوری و همکاران (۲۰۱۵) در تحقیقی به این

^۱ support vector machine (SVM)^۲ genetic programming^۳ regression kriging^۱ longitudinal dispersion coefficient (LDC)^۲ artificial neural network (ANN)^۳ adaptive neuro-fuzzy inference system (ANFIS)

شبکه ی عصبی مصنوعی

با توجه به مراجع کافی برای ANN (هیکین، ۱۹۹۹؛ مایر و دندی، ۲۰۰۰)، خوانندگان مقاله به مراجع ذکر شده ارجاع داده می‌شوند و در این تحقیق با توجه به حجم مطالب بررسی شده تنها به مطالب ضروری و مرتبط با کار انجام شده متناسب با موضوع مقاله پرداخته می‌شود. در استفاده از ANN مشکلی که همواره وجود دارد مشکل فوق برازشی^۲ یا کم برازشی^۴ شبکه است که در این حالت آموزش شبکه به خوبی انجام شده ولی شبکه از جامعیت مناسب برخوردار نیست؛ یعنی نتایج به دست آمده برای مراحل آموزش و آزمون شبکه تفاوت معناداری با یکدیگر دارند. برای رفع این مشکل از روش الگوریتم توقف آموزش^۵ استفاده شده است (کولیبالی و همکاران، ۲۰۰۰).

بر مبنای تحقیقات نظری و عملی انجام شده مشخص گردیده که شبکه با یک لایه پنهان به خوبی قادر به پیش‌بینی هر پارامتر غیرخطی و پیچیده‌ای است (هورنیک و همکاران، ۱۹۸۹؛ نوری و همکاران، ۲۰۱۰؛ نوری و همکاران، ۲۰۱۳). بنابراین ANN استفاده شده در این پژوهش از نوع سه لایه پس‌خور با الگوریتم پس‌انتشار خطا بوده و توابع انتقال در لایه‌ی پنهان و خروجی شبکه به ترتیب توابع سیگموئیدی و خطی انتخاب شدند. محدوده‌ی اختیارات محققین در استفاده از توابع آموزش جهت بهینه کردن وزن‌ها و بایاس شبکه با توجه به توابع آموزش فراوانی که معرفی شده‌اند، متنوع و زیاد است. این توابع به دو دسته کلی تقسیم می‌شود: دسته اول از روش‌های اکتشافی^۶ استفاده می‌کنند که معروفترین تابع آن Trainrp است. دسته دوم از روش‌های بهینه‌سازی عددی استاندارد^۷ بهره می‌برند که معروفترین تابع آن الگوریتم CG است. در این تحقیق از هر دو رویکرد ذکر شده برای آموزش مدل ANN جهت تخمین LDC استفاده شده است. در این راستا برای اولین بار با استفاده از توابع CG و Trainrp بهترین مدل ANN برای پیش‌بینی LDC روده‌ها توسعه داده شد.

سوال پاسخ دادند که استفاده از مدل‌های ANN، ANFIS و SVM در تخمین LDC تا چه حد قابل اعتمادند. به هر حال با مرور مطالعات قبلی و نتایج آنها مشخص می‌شود که به دلیل پیچیدگی رفتار LDC در شرایط مختلف، مدل‌های تجربی استفاده شده به‌وسیله برخی محققین (فیشور، ۱۹۷۵؛ سنو و چونگ، ۱۹۹۸؛ دنگ و همکاران، ۲۰۰۱؛ قیاسی و همکاران، ۲۰۱۷) قادر به تخمین دقیق این پارامتر نیستند (نوری و همکاران، ۲۰۰۹؛ نوری و همکاران، ۲۰۱۵). از طرفی نیز مطالعات انجام شده به‌وسیله مدل‌های پیشرفته‌تر نظیر ANN، محدود به بکارگیری نوع خاصی از تابع آموزش معرفی شده به‌وسیله لونبرگ (۱۹۴۴) و مارکویت (۱۹۶۳) برای بهینه نمودن وزن‌ها و بایاس شبکه است و عملکرد دیگر توابع آموزش برای بهینه نمودن پارامترهای مدل ANN در تحقیقات مرتبط با منابع آب کمتر گزارش گردیده است. بنابراین در راستای مطالب ذکر شده هدف اصلی این تحقیق امکان‌سنجی به‌کارگیری توابع آموزشی گردایان مزدوج^۱ (CG)، و همچنین تابع پس‌انتشار ارتجاعی^۲ (Trainrp) برای بهینه نمودن پارامترهای مدل ANN جهت دستیابی به مدل با دقت مطلوب تخمین LDC در روده‌های طبیعی قرار داده شد.

مواد و روشها

اطلاعات مسئله

در این تحقیق با جمع‌آوری مجموعه‌ای وسیع از داده‌های محققین مختلف اقدام به آموزش و آزمون ANN به منظور دستیابی به مدل مناسبی برای پیش‌بینی LDC شده است. ورودی‌ها به ANN ترکیبی از مشخصات هندسی و جریان در رودخانه‌ها شامل عرض رودخانه (W)، عمق رودخانه (H)، سرعت جریان (U) و سرعت برشی (u^*) جریان هستند و خروجی مدل نیز LDC است. این مشخصات شامل ۱۰۰ داده رودخانه‌های آمریکا هستند که در تحقیق اکبرزاده و همکاران (۱۳۸۹) با ذکر جزئیات آورده شده‌اند. این داده‌ها و نحوه تغییرات آنها در شکل شماره (۱) نشان داده شده‌اند.

⁵ stop training algorithm (STA)

⁶ heuristic

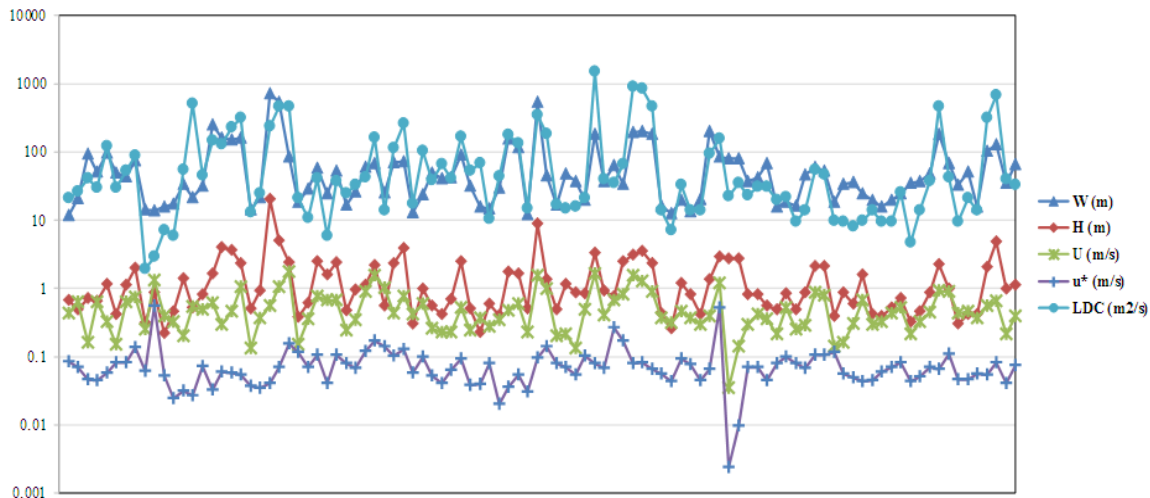
⁷ standard numerical optimization

¹ conjugate gradient (CG)

² resilient backpropagation

³ over-fitting

⁴ under-fitting



شکل ۱- نمودار ۱۰۰ داده استفاده شده در این تحقیق در مقیاس لگاریتمی

هستند. در ادامه، راستای جستجو به گونه‌ای تعیین می‌شود که به صورت مزدوج با راستای قبلی جستجو باشد. روند معمول برای تعیین راستای جدید جستجو، ترکیب مسیر بیشترین شیب جدید با مسیر جستجوی قبلی است که براساس روش محاسبه‌ی β_k ، روش‌های مختلفی برای الگوریتم گرادیان مزدوج ایجاد شده‌اند:

$$p_k = -g_k + \beta_k p_{k-1} \quad (3)$$

تابع آموزش فلچر-ریوس^۱ (Traincgf)

برای الگوریتم فلچر-ریوس فرایند به روزرسانی براساس روش زیر انجام می‌شود. این عبارت، نسبت توان دوم معیار گرادیان فعلی به توان دوم معیار گرادیان در مرحله‌ی قبلی است (فلچر و ریوس، ۱۹۶۴؛ هاگان و همکاران، ۱۹۹۶):

$$\beta_k = \frac{\frac{T}{g_k} g_k}{\frac{T}{g_{k-1}} g_{k-1}} \quad (4)$$

تابع آموزش پولاک-ریبره^۲ (Traincgp)

یکی دیگر از الگوریتم‌های CG، الگوریتم پولاک و ریبره است. برای تابع آموزش پولاک-ریبره، ضریب ثابت β_k از روش زیر محاسبه می‌شود. این عبارت ضرب داخلی تغییر قبلی در گرادیان با گرادیان فعلی تقسیم بر مربع معیار گرادیان در مرحله قبل است (فلچر و ریوس، ۱۹۶۴؛ هاگان و همکاران، ۱۹۹۶).

الگوریتم‌های CG

الگوریتم پایه پس‌انتشار، وزن‌ها را در جهتی که بیشترین شیب وجود دارد تنظیم می‌کند (معکوس گرادیان)، جهتی که در آن راستا تابع هدف (کمینه شدن خطا) به سریع‌ترین شکل ممکن نزول نماید. واضح است که اگرچه تابع هدف به سریع‌ترین شکل ممکن در جهت عکس گرادیان نزول می‌کند، اما این امر ضامنی برای ایجاد یک همگرایی سریع نیست. در الگوریتم‌های CG، جستجو در جهت‌های مزدوج، که معمولاً باعث ایجاد همگرایی سریعتر نسبت به جهت بیشترین شیب می‌شود، صورت می‌گیرد. همه‌ی روش‌های CG کارشان را در اولین مرحله براساس جستجو در جهت بیشترین شیب (معکوس گرادیان) آغاز می‌کنند، که در آن p_0 و g_0 به ترتیب بردارهای ورودی و گرادیان اولیه هستند.

$$p_0 = -g_0 \quad (1)$$

در ادامه برای مشخص کردن فاصله بهینه، مطابق رابطه‌ی زیر یک جستجوی خطی انجام می‌شود:

$$x_{k+1} = x_k + \alpha_k p_k \quad (2)$$

که در آن x_k و x_{k+1} به ترتیب بردارهای وزن و بایاس شبکه در تکرار k و $k+1$ هستند. α_k و p_k نیز به ترتیب برابر نرخ یادگیری و بردار ورودی به شبکه در تکرار k

² polak-ribière

¹ fletcher-reeves

کردن طول گامی که مقدار تابع هدف را کمینه کند، جستجویی در طول جهت گرادیان مزدوج انجام می شود. پنج تابع جستجوی متفاوت در جعبه ی ابزار ANN (نرم افزار MATLAB) وجود دارند. هر کدام از این توابع جستجو را می توان با یک تابع آموزش به کار برد. برخی از توابع جستجو با برخی از توابع آموزش دارای همخوانی بیشتری هستند، اگرچه انتخاب بهینه را می توان براساس کاربردهای خاص انجام داد. اغلب پیش بینی اینکه کدام یک از توابع جستجو جواب بهتری خواهند داد کار سختی است و تعیین آن از طریق سعی و خطا امکان پذیر است. در این تحقیق از سه تابع جستجو در استفاده همزمان از ۴ تابع آموزش دسته CG ذکر شده به شرح زیر استفاده شده است.

• تابع جستجوی برنت (Searchbre)

جستجوی برنت یک جستجوی خطی است که ترکیبی^۳ از جستجوی ناحیه ی طلایی (هاگان و همکاران، ۱۹۹۶) و یک درون یابی درجه ی دوم^۴ است (برنت، ۱۹۷۳). روش های مقایسه توابعی همانند جستجوی ناحیه طلایی دارای همگرایی درجه یک است؛ درحالی که درون یابی چند جمله ای دارای نرخ مجانبی است که سرعتی بیش از توابع خطی دارد. از سوی دیگر، نرخ همگرایی برای الگوریتم ناحیه ی طلایی هنگامی آغاز می شود که الگوریتم مقداردهی اولیه می شود، درحالی که رفتار مجانب روش های درون یابی چند جمله ای ممکن است بعد از چندین گردش به روز شود. تلاش روش جستجوی برنت، برای ترکیب خصوصیات خوب این دو روش است. برای جستجوی برنت، کار با همان بازه عدم اطمینان به کار رفته در روش جستجوی طلایی آغاز می گردد اما چندین نقطه ی اضافی نیز محاسبه می شوند. سپس یک تابع درجه ۲ به این نقاط برازش می شود و کمینه این تابع محاسبه می گردد. اگر این کمینه در داخل بازه ی مناسب عدم اطمینان باشد، در مرحله ی بعدی جستجو استفاده گشته و یک تخمین درجه ی ۲ جدید انجام می پذیرد. اگر کمینه خارج از ناحیه ی عدم اطمینان شناخته شده قرار گرفت، یک مرحله از جستجوی ناحیه ی طلایی انجام می پذیرد. این الگوریتم دارای این مزیت است که به محاسبه مشتق نیازی ندارد.

$$\beta_k = \frac{T g_k g_k}{T g_{k-1} - T g_{k-1}} \quad (5)$$

تابع آموزش پاول-بیل (Traincgb)

در همه الگوریتم های CG، جهت جستجو به صورت متناوب به مقدار منفی گرادیان تنظیم می شود. نقطه ی استاندارد برای تنظیم مجدد هنگامی است که تعداد گردش ها برابر، تعداد پارامترهای شبکه (وزن ها و بایاس ها) باشند. اما روش های دیگری نیز برای تنظیم مجدد وجود دارند که می توانند بازدهی آموزش را افزایش دهند. یکی از این روش ها به وسیله پاول (۱۹۷۷) بر مبنای روشی که پیشتر به وسیله بیل (۱۹۷۲) ارائه شده بود، پیشنهاد شده است. این روش، شروع مجدد را هنگامی انجام می دهد که یک تعامد کوچک بین مقدار فعلی گرادیان و مقدار قبلی گرادیان باقی مانده باشد. این امر با استفاده از نامساوی زیر آزموده می شود. اگر این شرط ارضا شد، جهت جستجو مجدداً به منفی گرادیان تنظیم می شود.

$$\left| g_{k-1}^T g_k \right| \geq 0.2 \|g_k\|^2 \quad (6)$$

تابع آموزش گرادیان مزدوج مقیاس دار ۲ (Trainscg)

هرکدام از روش های CG که تاکنون بحث شده اند به یک الگوریتم جستجوی خطی در هر گردش نیازمند است. از آنجا که پاسخ شبکه به همه ی ورودی های آموزش باید چندین بار محاسبه شوند، این جستجوی خطی به لحاظ محاسباتی دارای بار زیادی است. الگوریتم گرادیان مزدوج مقیاس دار (SCG) که به وسیله مولر (۱۹۹۳) طراحی شده است به گونه ای است که از محاسبات جستجوی خطی پرهیز می کند. این الگوریتم رویکرد ناحیه اعتماد (به کار گرفته شده در الگوریتم لونبرگ (۱۹۴۴) - مارکواد (۱۹۶۳) را با رویکرد گرادیان مزدوج ترکیب می کند. در بیشتر الگوریتم های آموزش استفاده شده جهت بهینه کردن وزن ها و بایاس شبکه، پارامتر نرخ آموزش برای تعیین کردن طول گام استفاده گردید، اما در الگوریتم های گرادیان مزدوج، مقدار طول گام در هر تکرار به صورت خودکار تنظیم می شود. به منظور تعیین

³ hybrid

⁴ quadratic interpolation

¹ powell-beale

² scaled conjugate gradient

بهینه‌ی خودشان فاصله داشته باشند. بنابراین هدف الگوریتم آموزش تابع Trainrp حذف کردن این تأثیرات مضر از بزرگی مشتقات جزئی است. در این روش، تنها، نشانه‌ای از مشتق برای تعیین مسیر به‌روزرسانی و وزن‌ها به کار می‌رود و مقدار مشتق تأثیری بر به‌روزرسانی وزن ندارد و اندازه تغییرات وزن به‌وسیله یک مقدار به‌روزرسانی جداگانه تعیین می‌گردد.

معیارهای استفاده شده جهت ارزیابی نتایج

برای مقایسه نتایج خروجی شبکه، ۳ معیار آماری متداول استفاده شد که عبارتند از: ضریب تعیین R^2 ، میانگین خطای مطلق (MAE)^۴ و ریشه‌ی میانگین مربعات خطا (RMSE)^۵. بهترین عملکرد مدل بر اساس معیار R^2 مقدار یک را به همراه دارد و برای سایر معیارها صفر است. همچنین در انتها برای مقایسه بهترین نتیجه توابع، از آماره نسبت تفاوت توسعه یافته^۶ (DDR) به‌وسیله نوری و همکاران (۲۰۱۰) استفاده گردید. در نمودار این آماره، هرچه تاج منحنی مقدار بزرگتری را داشته باشد، مدل ارزیابی شده با عملکرد بهتری همراه است.

بحث و نتایج

توسعه‌ی شبکه عصبی

هر کدام از متغیرها قبل از ورود به شبکه عصبی لازم است استاندارد شوند که در این تحقیق از رابطه‌ی (۷) استفاده شد.

$$x = \frac{2(x - x_{\min})}{(x_{\max} - x_{\min})} - 1 \quad (7)$$

در رابطه‌ی (۷) x_{\max} و x_{\min} به ترتیب حداقل و حداکثر مقادیر در مجموعه داده‌های آموزش و آزمون مدل هستند. متغیر x نیز مقادیر استاندارد شده متغیرها است. با استفاده از این رابطه، محدوده‌ی هر متغیر بین -۱ تا +۱ قرار می‌گیرد. بعد از استاندارد کردن متغیرها با تغییر تعداد نرون‌ها در لایه‌ی پنهان شبکه، مدل‌های متفاوتی، با استفاده از هر کدام از توابع آموزشی استفاده شده در این مقاله، به منظور پیش‌بینی LDC ایجاد شد. برای انتخاب بهترین تابع جستجوی مناسب برای هر کدام از توابع

• تابع جستجوی ترکیبی نیمه-مکعبی^۱ (Searchhyb) همانند روش برنت، روش تابع جستجوی ترکیبی نیمه-مکعبی یک روش ترکیبی است. این روش ترکیبی از الگوریتم نصف کردن و درون‌یابی درجه‌ی ۳ (مکعبی) است. برای الگوریتم نصف کردن، یک نقطه در درون بازه‌ی عدم اطمینان انتخاب می‌شود و برای آن، تابع هدف و مقادیر مشتقات محاسبه می‌گردد. براساس این اطلاعات، از نصف بازه‌ی عدم اطمینان صرف‌نظر می‌شود. در الگوریتم ترکیبی، با استفاده از مقادیر بازه و مشتقات آن در نقاط ابتدا و انتها، یک تابع برای درون‌یابی درجه‌ی ۳ به دست می‌آید. اگر مقدار به دست آمده براساس درون‌یابی درجه‌ی ۳ درون بازه‌ی عدم اطمینان باشد، از آن برای کاهش بازه جستجو استفاده شده، و در غیر این صورت، یک گام با استفاده از الگوریتم نصف کردن پیش‌روی صورت می‌گیرد. برای اطلاعات بیشتر در این زمینه می‌توان به اسکالس (۱۹۸۵) مراجعه نمود.

• تابع جستجوی چارالامبوس^۲ (Searchcha) روش چارالامبوس، Searchcha، برای ترکیب با یک الگوریتم CG برای آموزش شبکه‌ی عصبی طراحی شده است. همانند دو روش قبلی، این روش نیز یک جستجوی ترکیبی است. این روش از ترکیب درون‌یابی درجه‌ی ۳ با یک نوع الگوریتم تکه تکه کردن تشکیل شده است (چارالامبوس، ۱۹۹۲).

تابع Trainrp

شبکه‌های چندلایه‌ای معمولاً در لایه‌های پنهان خود از توابع انتقالی سیگموئید استفاده می‌کنند. این توابع اغلب "فشرده کننده"^۳ نامیده می‌شوند زیرا داده‌های ورودی با بازه‌ی بی‌نهایت را به صورت خروجی با بازه‌ی محدود، فشرده می‌کنند. توابع سیگموئید، با این حقیقت که در مواقعی که مقادیر ورودی بزرگ شوند، شیب آن‌ها نزدیک صفر می‌گردد، شناخته شده‌اند. این موضوع زمانی که بخواهیم در آموزش یک شبکه چندلایه‌ای با تابع سیگموئید از بیشترین کاهش شیب استفاده کنیم، مشکل ایجاد می‌نماید زیرا با توجه مقدار بسیار کوچک متغیر، تغییرات کوچکی در وزن‌ها و بایاس‌ها ایجاد می‌کند، در حالی که ممکن است این وزن‌ها و بایاس‌ها از مقادیر

⁴ mean absolute error

⁵ root mean square error

⁶ developed discrepancy ratio

¹ hybrid bisection-cubic

² charalambous

³ squashing

در مرحله آزمون با اختلاف زیاد کمترین خطا را شامل می‌شود) نسبت به بقیه مدل‌ها از عملکرد مطلوب‌تری برخوردار است. پس این ساختار به عنوان ساختار بهینه در این مرحله از تحقیق انتخاب می‌شود.

آموزش شبکه با تابع آموزش پاول-بیل

مشابه مرحله‌ی قبل مدل‌های متفاوتی برای این تابع آموزش با تغییر در نوع تابع جستجو ایجاد شد. همچنین، تعداد نرون‌ها در لایه پنهان ۴، ۶، ۸ و ۱۰ نرون برای هر مدل انتخاب شد. نتایج مراحل آموزش و آزمون هر مدل در جدول ۳ آمده‌اند. در این مرحله با توجه به نتایج به‌دست آمده در مجموع مراحل آموزش و آزمون، مشخص است که مدل با تابع جستجوی Searchhyb دارای ۸ لایه‌ی پنهان در این مرحله به عنوان ساختار بهینه انتخاب می‌شود.

آموزش شبکه با تابع آموزش SCG

در این مرحله نیز مدل‌های متفاوتی برای تابع آموزش SCG با تغییر در نوع تابع جستجو ایجاد شدند. همچنین، تعداد نرون‌ها در لایه پنهان ۴، ۶، ۸ و ۱۰ نرون برای هر مدل انتخاب شد. نتایج مراحل آموزش و آزمون هر مدل در جدول ۴ آمده‌اند. با توجه به آماره‌ی R^2 در مراحل آموزش و آزمون، تابع Searchhyb دارای ۶ لایه پنهان در این مرحله به عنوان ساختار بهینه انتخاب می‌شود.

تابع Trainrp

نتایج به‌دست آمده از این تابع برای مدل‌های با ۴، ۶، ۸ و ۱۰ نرون در لایه‌ی پنهان به ترتیب با ساختارهای (۴-۴-۱)، (۴-۶-۱)، (۴-۸-۱) و (۴-۱۰-۱) در جدول ۵ نشان داده شده‌اند. با توجه به این جدول مشخص می‌شود که مدل با ۴، ۸ و ۱۰ نرون در لایه‌ی پنهان دارای نتایجی بهتر از مدل با ۶ نرون در لایه‌ی پنهان هستند. همانطور که از جدول ۵ مشخص است، در این روش مدل ایجاد شده با ساختار (۴-۴-۱) در مجموع عملکرد به مراتب بهتری را نسبت به سایر ساختارها دارد.

انتخاب بهترین مدل

در مراحل قبل مدل‌های مناسبی با استفاده از چهار نوع تابع آموزش CG و تابع آموزش Trainrp آموزش داده

آموزش نیز از روش سعی و خطا استفاده گردید. بدین ترتیب که با تغییر توابع جستجو در هر تابع آموزش استفاده شده، بهترین تابع جستجو برای آن تابع آموزش مشخص، معین گردید. همچنین با تغییرات مناسب در کد مربوط به ANN در نرم‌افزار MATLAB، برنامه‌ریزی به طریقی انجام گرفت که برای هر ساختار از شبکه، ۱۰ بار شبکه اجرا شود و در نهایت نیز بهترین نتایج حاصل را به عنوان خروجی برنامه ذخیره نماید.

آموزش شبکه با تابع آموزش فلچر-ریوس

همان‌طور که قبلاً ذکر شد، در الگوریتم‌های CG، مقدار طول گام در هر تکرار به صورت خودکار تنظیم می‌شود. به منظور تعیین کردن طول گامی که مقدار تابع هدف را کمینه می‌کند، جستجویی در طول جهت CG انجام می‌شود. در این مرحله از تحقیق با استفاده از سعی و خطا و انتخاب هر یک از سه تابع جستجوی معرفی شده در قبل به عنوان توابع جستجو برای تابع آموزش فلچر-ریوس، مدل‌های متفاوتی با تغییر در تعداد نرون‌های لایه‌ی پنهان شبکه ایجاد شد. با توجه به تعداد اندک ورودی‌ها و به منظور جلوگیری از پیچیده شدن ساختار شبکه، تعداد نرون‌ها در لایه‌ی پنهان ۴، ۶، ۸ و ۱۰ نرون برای هر مدل انتخاب شدند. نتایج مراحل آموزش و آزمون هر مدل در جدول ۱ آمده‌اند. با توجه به نتایج به دست آمده از این تابع آموزش، مشخص است که تغییر تابع جستجو، تغییرات قابل توجهی در مقدار آماره‌های بررسی شده در این مقاله ندارد. به هر حال، با توجه به جدول ۱، مشخص می‌شود که تابع جستجوی چارالامبوس (Searchcha) با ۸ نرون در لایه‌ی پنهان از نتایج مطلوب‌تری نسبت به بقیه مدل‌ها برخوردار است.

آموزش شبکه با تابع آموزش پولاک-ریبره

مشابه مرحله قبل مدل‌های متفاوتی برای این تابع آموزش با تغییر در نوع تابع جستجو و همچنین تعداد نرون‌ها در لایه‌ی پنهان شبکه ایجاد شد که نتایج مراحل آموزش و آزمون هر مدل در جدول ۲ آمده‌اند. در این مرحله نیز با توجه به نتایج به‌دست آمده، تابع جستجوی چارالامبوس با ۸ نرون در لایه‌ی پنهان شبکه با توجه به جمع‌بندی مقادیر R^2 ، RMSE و MAE در مراحل آموزش و آزمون (در مرحله آموزش دارای رتبه‌ی دوم است ولی

و ارزیابی شدند. در هر مرحله نیز بهترین مدل انتخاب شد که نتایج این مدل‌های بهینه در جدول ۶ آمده‌اند.

جدول ۱- نتایج حاصل از مرحله‌ی آموزش و آزمون شبکه‌ی عصبی برای تابع آموزش فلچر-ریوس

آزمون			آموزش			نرون	تابع جستجو	تابع آموزش
MAE	RMSE	R ²	MAE	RMSE	R ²			
۸۶	۱۵۶	۰/۷۰	۸۶	۱۴۳	۰/۶۸	۴	Searchbre	
۸۰	۱۶۱	۰/۷۸	۹۳	۲۰۶	۰/۸۰	۶		
۶۰	۱۰۶	۰/۹۰	۳۹	۷۵	۰/۹۱	۸		
۶۹	۱۲۰	۰/۷۷	۶۴	۱۱۶	۰/۸۰	۱۰		
۴۴	۸۵	۰/۸۵	۵۰	۱۱۰	۰/۸۲	۴	Searchcha	Traincgrf
۵۱	۱۰۴	۰/۹۳	۴۳	۷۶	۰/۹۲	۶		
۴۱	۷۸	۰/۹۴	۴۱	۷۰	۰/۹۲	۸		
۸۰	۱۶۵	۰/۶۳	۵۵	۱۱۴	۰/۷۹	۱۰		
۶۹	۱۲۳	۰/۶۶	۷۳	۱۴۲	۰/۶۶	۴	Searchhyb	
۵۹	۱۲۶	۰/۷۲	۵۷	۱۰۸	۰/۸۲	۶		
۴۸	۱۰۲	۰/۸۰	۵۸	۱۱۰	۰/۸۰	۸		
۸۱	۱۵۶	۰/۷۸	۵۹	۱۰۸	۰/۸۱	۱۰		

جدول ۲- نتایج حاصل از مرحله‌ی آموزش و آزمون شبکه‌ی عصبی برای تابع آموزش پولاک-ریبره

آزمون			آموزش			نرون	تابع جستجو	تابع آموزش
MAE	RMSE	R ²	MAE	RMSE	R ²			
۷۸	۹۸	۰/۸۴	۷۷	۱۱۲	۰/۸۱	۴	Searchbre	
۵۰	۸۷	۰/۸۳	۶۲	۱۰۹	۰/۸۲	۶		
۵۵	۱۲۰	۰/۷۲	۵۷	۱۱۱	۰/۸۰	۸		
۵۳	۱۱۲	۰/۹۲	۳۹	۷۴	۰/۹۱	۱۰		
۶۱	۸۵	۰/۸۸	۵۹	۹۶	۰/۸۵	۴	Searchcha	Traincgp
۶۷	۱۳۰	۰/۶۸	۶۳	۱۲۹	۰/۷۲	۶		
۴۰	۶۴	۰/۹۲	۴۸	۸۷	۰/۸۹	۸		
۷۸	۱۳۸	۰/۷۹	۵۴	۱۰۵	۰/۸۱	۱۰		
۶۸	۱۴۳	۰/۷۳	۴۷	۱۰۰	۰/۸۳	۴	Searchhyb	
۷۲	۱۰۰	۰/۷۷	۷۶	۱۳۹	۰/۷۵	۶		
۷۲	۱۴۵	۰/۷۳	۶۲	۱۱۶	۰/۷۷	۸		
۹۹	۱۷۶	۰/۶۶	۸۶	۱۳۲	۰/۷۴	۱۰		

جدول ۳- نتایج حاصل از مرحله‌ی آموزش و آزمون شبکه‌ی عصبی برای تابع آموزش پاول-بیل

آزمون			آموزش			نرون	تابع جستجو	تابع آموزش
MAE	RMSE	R ²	MAE	RMSE	R ²			
۵۲	۱۲۳	۰/۶۷	۳۱	۵۱	۰/۹۶	۴	Searchbre	
۴۸	۱۳۴	۰/۶۹	۵۵	۱۰۴	۰/۸۲	۶		
۴۹	۱۲۷	۰/۷۰	۴۷	۸۰	۰/۹۰	۸		
۵۵	۱۱۸	۰/۷۲	۵۵	۱۰۷	۰/۸۱	۱۰		
۵۳	۱۲۷	۰/۷۶	۵۶	۱۰۹	۰/۸۰	۴	Searchcha	Traincgb
۵۵	۹۷	۰/۸۰	۶۱	۱۱۴	۰/۸۰	۶		
۷۸	۱۱۴	۰/۹۰	۸۶	۱۵۶	۰/۵۹	۸		
۵۷	۱۳۴	۰/۷۴	۴۰	۸۰	۰/۸۹	۱۰		
۵۲	۱۵۴	۰/۵۹	۴۲	۶۲	۰/۹۴	۴	Searchhyb	
۷۵	۸۴	۰/۹۲	۷۲	۱۳۳	۰/۷۲	۶		
۵۰	۷۳	۰/۹۰	۴۵	۷۲	۰/۹۲	۸		
۶۵	۹۶	۰/۸۴	۳۷	۶۷	۰/۹۲	۱۰		

جدول ۴- نتایج حاصل از مرحله‌ی آموزش و آزمون شبکه‌ی عصبی برای تابع آموزش گرادبان مزدوج مقیاس‌دار

آزمون			آموزش			نرون	تابع جستجو	تابع آموزش
MAE	RMSE	R ²	MAE	RMSE	R ²			
۶۸	۱۳۳	۰/۶۷	۶۵	۱۴۱	۰/۶۷	۴	Searchbre	
۷۱	۱۳۴	۰/۶۹	۶۲	۱۳۱	۰/۷۱	۶		
۶۴	۱۲۷	۰/۷۰	۶۳	۱۱۹	۰/۷۶	۸		
۶۵	۱۱۸	۰/۷۲	۶۷	۱۳۷	۰/۶۹	۱۰		
۷۳	۱۲۷	۰/۷۶	۵۹	۱۱۹	۰/۷۷	۴	Searchcha	Trainscg
۶۰	۹۷	۰/۸۰	۶۱	۱۰۹	۰/۸۷	۶		
۵۹	۱۱۴	۰/۹۰	۴۲	۰/۸۰	۰/۸۹	۸		
۶۸	۱۳۴	۰/۷۴	۵۲	۱۰۵	۰/۸۱	۱۰		
۱۳۵	۱۵۴	۰/۵۹	۱۳۷	۱۷۸	۰/۶۱	۴	Searchhyb	
۴۸	۸۴	۰/۹۲	۴۱	۶۴	۰/۹۳	۶		
۵۴	۷۳	۰/۹۰	۶۰	۱۱۱	۰/۸۸	۸		
۶۷	۹۶	۰/۸۴	۷۵	۱۱۴	۰/۸۳	۱۰		

جدول ۵- نتایج به‌دست آمده از مدل‌های مختلف ساخته شده به‌وسیله‌ی شبکه‌ی عصبی تابع Trainrp

آزمون			آموزش			ساختار شبکه
MAE	RMSE	R ²	MAE	RMSE	R ²	
۳۷	۷۰	۰/۹۰	۴۱	۸۹	۰/۸۷	(۴-۴-۱)
۸۷	۱۳۹	۰/۷۰	۸۶	۱۳۰	۰/۷۶	(۴-۶-۱)
۵۷	۹۸	۰/۸۸	۵۰	۸۴	۰/۸۸	(۴-۸-۱)
۶۴	۱۰۹	۰/۸۷	۵۵	۹۶	۰/۸۵	(۴-۱۰-۱)

جدول ۶- نتایج حاصل از مرحله‌ی آموزش و آزمون شبکه‌ی عصبی برای مدل‌های منتخب

آزمون			آموزش			نرون	تابع جستجو	تابع آموزش
MAE	RMSE	R ²	MAE	RMSE	R ²			
۴۸	۸۴	۰/۹۲	۴۱	۶۴	۰/۹۳	۶	Searchhyb	Trainscg
۴۱	۷۸	۰/۹۴	۴۱	۷۰	۰/۹۲	۸	Searchcha	Traincgf
۷۸	۱۳۸	۰/۷۹	۵۴	۱۰۵	۰/۸۱	۱۰	Searchcha	Traincgp
۵۰	۷۳	۰/۹۰	۴۵	۷۲	۰/۹۲	۸	Searchhyb	Traincgb
۳۷	۷۰	۰/۹۰	۴۱	۸۹	۰/۸۷	۴	---	Trainrp (4-4-1)

خطای زیادی دارد. با مقایسه بین توابع Traincgf و Trainrp، مشخص می‌شود که در مرحله‌ی آموزش، تابع Traincgf بهتر عمل می‌کند و در مرحله‌ی آزمون، تابع Trainrp خطای کمتری را دارد. بنابراین برای بررسی بهتر عملکرد این دو تابع از آماره‌ی DDR استفاده گردید. همانطور که در معرفی آماره‌ی DDR ذکر شد، و با توجه به شکل ۲، تابع Trainrp از عملکرد به مراتب بهتری نسبت به تابع Traincgf برخوردار است. پس با توجه به

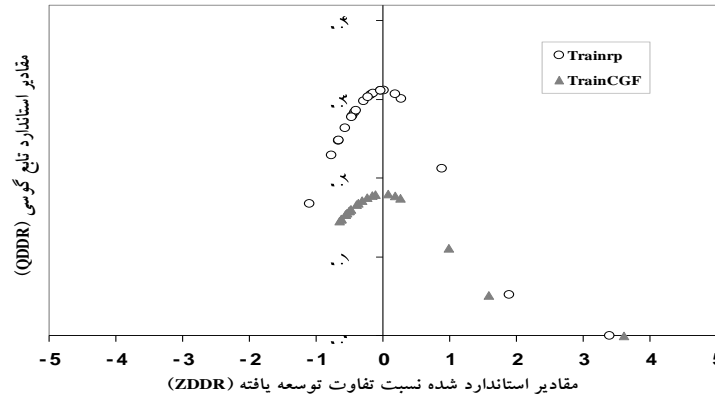
در این مرحله با مقایسه‌ی نتایج به‌دست آمده از این مدل‌ها، بهترین مدل به منظور پیش‌بینی LDC در رودخانه‌های طبیعی انتخاب می‌شود. همان‌طور که از جدول ۶ مشخص است، در بین بهترین مدل‌های الگوریتم CG، تابع Traincgf با ۸ نرون در لایه‌ی پنهان خود، عملکرد بهتری دارد. گرچه تابع Traincgb با ۴ نرون در لایه پنهان خود در مرحله‌ی آموزش خطای کمتری دارد، ولی این تابع در مرحله‌ی آزمون، با وجود ضریب R² بالا،

ریاحی مدوار (۲۰۰۹) دارای بهترین مقادیر R^2 هستند، ولی مقادیر بالای آماره‌های RMSE و MAE در این مدل‌ها باعث می‌شود که به دلیل خطای زیاد، نتوان در تخمین LDC به این مدل‌ها اطمینان کرد. دو مدل آدارش (۲۰۱۰) و الگوریتم Trainrp نیز مقادیر R^2 مناسبی را دارند، و علاوه بر این، مقادیر خطای این دو مدل نیز بسیار کمتر از دو مدل پارسایی و حقی‌آبی-الف (۲۰۱۳) و ریاحی‌مدوار (۲۰۰۹) است. در نتیجه می‌توان گفت که دو مدل آدارش (۲۰۱۰) و الگوریتم Trainrp در مقایسه با سایر مدل‌ها بهترین عملکرد را دارند. نتایج جدول ۷ بیانگر این نکته‌اند که هرچند مقدار RMSE الگوریتم Trainrp کمی بیشتر از مدل آدارش (۲۰۱۰) است، ولی مقدار MAE آن به مراتب بهتر از مدل آدارش (۲۰۱۰) است و الگوریتم Trainrp عملکرد بهتری را دارد. در نتیجه در مقایسه‌ای بین مدل‌های هوشمند استفاده شده به وسیله دیگر محققان برای تخمین LDC و نتایج این تحقیق، الگوریتم Trainrp به عنوان مناسب‌ترین مدل برای تخمین LDC در روده‌های طبیعی پیشنهاد می‌شود.

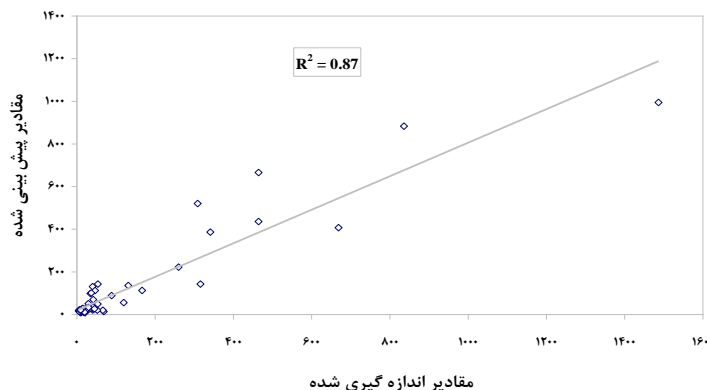
مطالب ذکر شده، مدل ایجاد شده Trainrp به‌وسیله مدل با ساختار (۴-۴-۱) به‌عنوان مدل برگزیده برای پیش‌بینی LDC در این تحقیق انتخاب می‌شود. نتایج مراحل آموزش و آزمون این مدل در شکل‌های ۳ تا ۶ آمده‌اند. با توجه به این شکل‌ها مشخص می‌شود که مدل مذکور در مرحله آزمون (شکل ۶) اگر چه تخمین LDC با مقادیر بالا را با دقت زیادی به خوبی انجام داده است، ولی برای LDC با مقادیر کم این مدل مقادیر را بیشتر از مقادیر واقعی پیش‌بینی نموده است.

مقایسه با نتایج دیگر تحقیقات

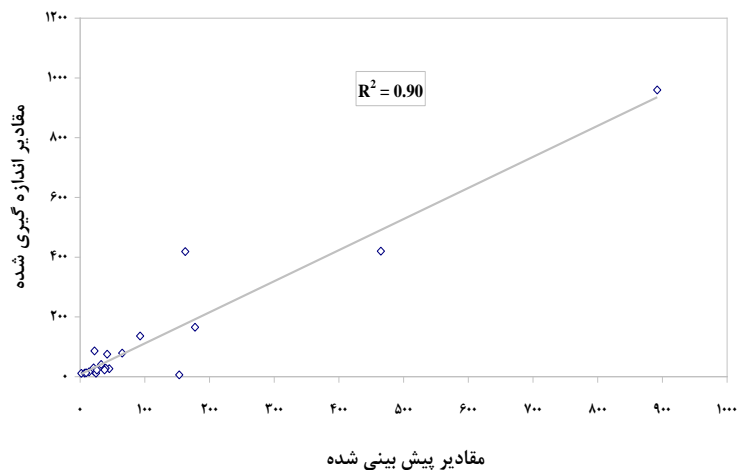
در پایان، مطابق جدول زیر مقایسه‌ای بین نتایج بهترین مدل این تحقیق با نتایج دیگر مطالعات انجام شده به‌وسیله مدل‌های هوشمند و تجربی که برای تخمین LDC تحقق یافته، صورت گرفته است. همان‌طور که از جدول ۷ مشخص است عملکرد مدل‌های تجربی تقریباً یکسان بوده و در مقایسه با مدل توسعه داده شده در این تحقیق مقدار خطای آنها بیشتر و مقدار R^2 آنها کمتر است. دو مدل پارسایی و حقی‌آبی-الف (۲۰۱۳) و



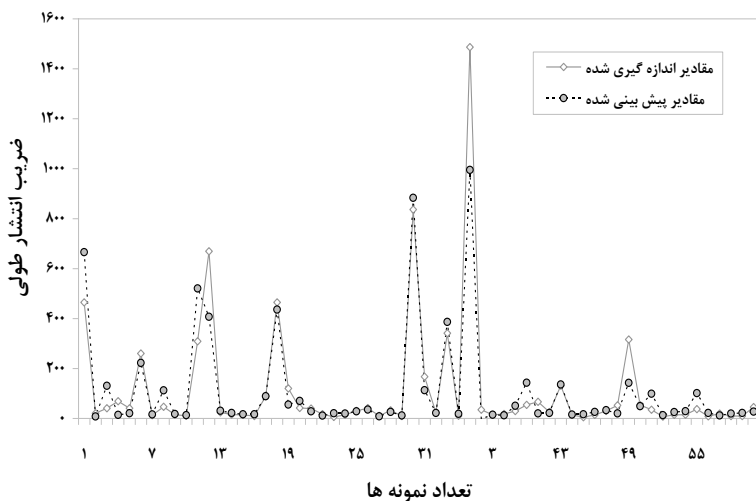
شکل ۲- نمودار آماری DDR برای مقایسه‌ی بین توابع Traincgf و Trainrp



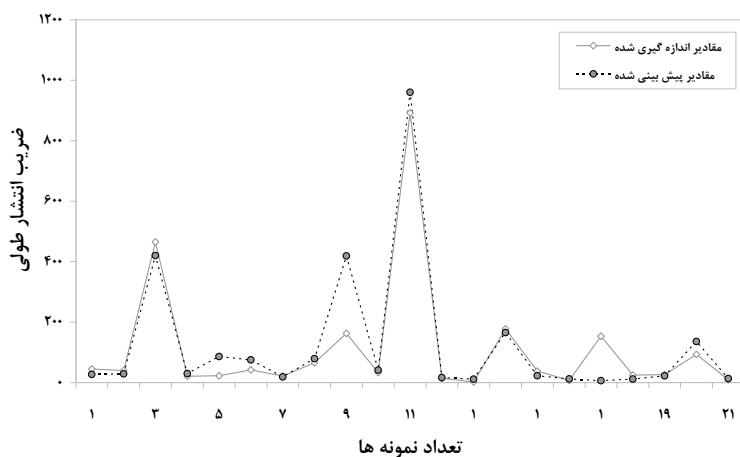
شکل ۳- نتایج مدل شبکه‌ی عصبی با ساختار (۴-۴-۱) در مرحله‌ی آموزش مدل



شکل ۴- نتایج مدل شبکه‌ی عصبی با ساختار (۴-۴-۱) در مرحله‌ی آزمون مدل



شکل ۵- نتایج مدل شبکه‌ی عصبی با ساختار (۴-۴-۱) در مرحله‌ی آموزش مدل



شکل ۶- نتایج مدل شبکه‌ی عصبی با ساختار (۴-۴-۱) در مرحله‌ی آزمون مدل

جدول ۷- مقایسه‌ی نتایج بهترین مدل‌های این تحقیق با دیگر مطالعات انجام شده به‌وسیله‌ی مدل‌های هوشمند و تجربی

MAE	RMSE	R ²	مدل
۴۸/۳۸	۸۴/۴۷	۰/۶۸	سئو و چئونگ (۱۹۹۸)
۴۱/۹۱	۸۲/۷۹	۰/۷۱	دنگ و همکاران (۲۰۰۱)
۴۵/۹۲	۹۷/۰۷	۰/۶۴	کاشفی پور و فالكونر (۲۰۰۲)
۴۶/۸۷	۹۲/۹۹	۰/۶۰	سهای و دوتا (۲۰۰۹)
۴۲/۴۲	۹۶/۷۴	۰/۶۸	ژنگ و هوآی (۲۰۱۴)
---	۱۳۰/۶۲	۰/۹۹	پارسایی و حقی‌آبی - الف (۲۰۱۳)
---	۱۰۲/۲۸	۰/۸۵	پارسایی و حقی‌آبی - ب (۲۰۱۳)
۱۸۵۱	۱۹۲/۹۱	۰/۶۹۰۶	تففور و همکاران (۲۰۰۵)
۹۵/۹۴	۲۳۰/۴۱	۰/۵۰۴۱	نوری و همکاران (۲۰۰۹)
۹۰/۲	۲۳۵/۳	۰/۵۳۲۹	نوری و همکاران (۲۰۰۹)
۱۰۴/۷۷	۱۸۷/۸	۰/۹۰۸۴	ریاحی‌مدوار (۲۰۰۹)
۵۴/۴۲	۶۰/۴۴	۰/۸۹۳۰	آدارش (۲۰۱۰)
۲۹/۵	۳۸/۶۶	۰/۸۱	توتمز (۲۰۱۳)
۳۷/۳۳	۶۹/۸۳	۰/۹۰	الگوریتم Trainrp

داده شده برای دیگر کاربران در بقیه نقاط دنیا موجود نیست. مهمترین نتایج به دست آمده از این پژوهش عبارت بودند از:

- در میان توابع آموزشی CG، تابع Traincgf دارای بهترین عملکرد است.
- در مقایسه دو رویکرد بهینه‌سازی عددی استاندارد و رویکرد اکتشافی، نتایج رویکرد بهینه‌سازی عددی استاندارد همبستگی بهتری با نتایج واقعی داشت، در حالی که عملکرد رویکرد اکتشافی در تخمین LDC بهتر بود و خطای کمتری را شامل می‌شد.
- تابع آموزشی Trainrp در مقایسه با توابع آموزشی CG، از عملکرد بهتری برخوردار بودند.
- بهترین مدل انتخاب شده در این تحقیق، عملکرد به مراتب بهتری نسبت به دیگر مطالعات انجام شده با مدل‌های هوشمند و تجربی را دارا بود.

منابع

- 1) Akbarzadeh, A., Noori, R., Farokhnia, A., Khakpour, A., Sabahi, M. S. 2010. Accuracy and Uncertainty Analysis of Intelligent Techniques for Predicting the Longitudinal Dispersion Coefficient in Rivers. J. of Water and Wastewater. 21(3): 99-107. (In Persian)
- 2) Adarsh, S. 2010. Prediction of longitudinal dispersion coefficient in natural channels using soft computing

نتیجه‌گیری

این مطالعه تلاشی در جهت دستیابی به مدل بهینه‌ی پیش‌بینی LDC در رودهای طبیعی است. بدین منظور در ابتدا با مرور برخی از مطالعات انجام گرفته در زمینه مزبور، این نکته محرز گشت که به دلیل پیچیدگی رفتار این پارامتر، مدل‌های کلاسیک ارائه شده قادر به پیش‌بینی دقیق آن نیستند، و از طرفی نیز استفاده از مدل‌های قویتر، همچون ANN در برخی مطالعات نیاز به بازبینی بیشتری دارد. به عنوان مثال مطالعات انجام شده به‌وسیله مدل‌های ANN، محدود به به‌کارگیری نوع خاصی از تابع آموزش معرفی شده به‌وسیله لونبرگ (۱۹۴۴) و مارکویت (۱۹۶۳) برای بهینه نمودن وزن‌ها و بایاس شبکه است و عملکرد دیگر توابع آموزش برای بهینه نمودن پارامترهای مدل ANN به ندرت در تحقیقات مرتبط با منابع آب گزارش گردیده است. بدین منظور، در این مقاله با توجه به ناتوانی مدل‌های رگرسیونی و همچنین امکان‌سنجی به‌کارگیری دیگر الگوریتم‌های آموزشی برای بهینه نمودن پارامترهای مدل ANN، مدل‌های متفاوتی از ANN با توابع متنوع آموزش برای پیش‌بینی LDC در رودخانه‌های طبیعی آموزش داده و آزمون شوند. البته این نکته نیز قابل ذکر است که روش استفاده شده در این تحقیق از روشهای جعبه سیاه است و متاسفانه در مقایسه با روشهای رگرسیونی کاربرپسند نبوده و امکان استفاده مستقیم از مدل توسعه

- linear regression model. Iranian Water Research Journal. 24(11): 97-108. (In Persian)
- 15) Givehch, M., Maghrebi, M. F., Abrishami, J. 2008. Application of Depth-Averaged Velocity Profile for Estimation of Longitudinal Dispersion in Rivers. J. of Water and Wastewater. 20(4): 91-96. (In Persian)
- 16) Haykin, S. 1994. Neural networks: a comprehensive foundation. Prentice Hall PTR.
- 17) Hornik, K., Stinchcombe, M., White, H. 1989. Multilayer feedforward networks are universal approximators. Neural networks 2.5: 359-366.
- 18) Huber, P. J. 1981. Robust statistics. John Wiley & Sons, Inc. New York.
- 19) Jalili-Ghazizade, M., Noori, R. 2008. Prediction of Municipal Solid Waste Generation by Use of Artificial Neural Network: A Case Study of Mashhad. International Journal of Environmental Research. 2: 22-31.
- 20) Kashefipour, M. S., and Falconer, R. A. 2002. Longitudinal dispersion coefficients in natural channels. Water Research. 36(6): 1596-1608.
- 21) Kashefipour, M. 2007. Prediction of longitudinal dispersion coefficient in natural rivers using artificial neural networks. Iranian Journal of Hydraulic. 3: 15-25. (In Persian)
- 22) Levenberg, K. 1944. A method for the solution of certain non-linear problems in least squares. Quarterly of applied mathematics. 2(2): 164-168.
- 23) Marquardt, D. W. 1963. An algorithm for least-squares estimation of nonlinear parameters. Journal of the society for Industrial and Applied Mathematics. 11(2): 431-441.
- 24) Maier, H. R., & Dandy, G. C. 2000. Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications. Environmental modelling & software. 15(1): 101-124.
- techniques. Scientia Iranica. Transaction A, Civil Engineering. 17(5): 363.
- 3) Beale, E. M. L. 1972. A derivation of conjugate gradients In F A. Lootsma (Ed), Numerical methods/or nonlinear optimization. London: Academic Press.
- 4) Brent, R. P. 1973. Algorithms for Minimization without Derivatives, Englewood Cliffs. NJ: Prentice-Hall.
- 5) Charalambous, C. 1992. Conjugate gradient algorithm for efficient training of artificial neural networks. IEEE Proceedings. 139(3): 301-310.
- 6) Chow, V. T. 1973. Open channel hydraulics. McGraw-Hill Company. pp.680.
- 7) Coulibaly, P., Anctil, F., & Bobee, B. 2000. Daily reservoir inflow forecasting using artificial neural networks with stopped training approach. Journal of Hydrology. 230(3-4): 244-257.
- 8) Deng, Z. Q., Singh, V. P., and Bengtsson, L. 2001. Longitudinal dispersion coefficient in straight rivers. Journal of hydraulic engineering. 127(11): 919-927.
- 9) Etemad-Shahidi, A., and Taghipour, M. 2012. Predicting Longitudinal Dispersion Coefficient in Natural Streams Using M5' Model Tree. Journal of hydraulic engineering. 138(6): 542-554.
- 10) Fischer, H. B., List, J. E., Koh, C. R., Imberger, J., & Brooks, N. H. 1979. Mixing in inland and coastal waters. Elsevier. New York. 104-138.
- 11) Fischer, H. B. 1975. Discussion of 'Simple method for predicting dispersion in streams,' by R. S. McQuivey and T. N. Keifer. Journal of the Environmental Engineering Division. 101(3): 453-455.
- 12) Fletcher, R., and Reeves, C. M. 1964. Function minimization by conjugate gradients. Computer Journal. 7: 149-154.
- 13) Hagan, M. T., Demuth, H. B., and Beale, M. H. 1996. Neural Network Design. Boston. MA: PWS Publishing, 1996.
- 14) Ghiasi, B., Noori, R., Karbasi, A. R., Deng, Z. 2017. Estimating Longitudinal Dispersion Coefficient in rivers using non-

- of Mashhad. *Environmental Progress and Sustainable Energy*. 28 (2): 249-258.
- 33) Noori, R., Kerachian, R., Khodadadi, A., and Shakibinia, A. 2007. Assessment of importance of water quality monitoring stations using principal component and factor analysis: A case study of the Karoon River. *Journal of Water and Wastewater*. 63: 60-69. (In Persian)
- 34) Parsaie, A. and Haghiahi, A. H. 2013. Evaluation of experimental provided equations and artificial intelligence models to compute the longitudinal dispersion of pollution in rivers. *Journal of Environmental Management and Planing*. 3(2): 57-71. (In Persian)
- 35) Powell, M. J. D. 1977. Restart procedures for the conjugate gradient method. *Mathematical Programming*. 12: 241-254.
- 36) Riahi-Madvar, H., Ayyoubzadeh, S. A., Khadangi, E., & Ebadzadeh, M. M. 2009. An expert system for predicting longitudinal dispersion coefficient in natural streams by using ANFIS. *Expert Systems with Applications*. 36(4): 8589-8596.
- 37) Riahi-madvar, H., and Ayyaoubzadeh, S. A. 2008. Estimating longitudinal dispersion coefficient of pollutants using adaptive neuro-fuzzy inference system. *Journal of Water and Wastewater*. 67: 34-46. (In Persian)
- 38) Rutherford, J. C. 1994. *River mixing*. John Wiley & Son Limited.
- 39) Scales, L. E. 1985. *Introduction to Non-Linear Optimization*. New York: Springer-Verlag.
- 40) Seo, I. W., Cheong, T. S. 1998. Predicting longitudinal dispersion coefficient in natural Streams. *Journal of hydraulic engineering*. 124(1): 25-32.
- 41) Toprak, Z. F., Sen, Z., and Savci, M. E. 2004. Longitudinal dispersion coefficients in natural channels. *Water Research*. 38: 3193-3143.
- 42) Tutmez, B., and Mehmet, Y. 2013. Regression Kriging Analysis for Longitudinal Dispersion Coefficient.
- 25) Moller, M. F. 1993. A scaled conjugate gradient algorithm for fast supervised learning. *Neural Networks*. 6: 525-533.
- 26) Noori, R., Ghiasi, B., Sheikhan, H., Adamowski, J. F. 2017. Estimation of the Dispersion Coefficient in Natural Rivers Using a Granular Computing Model. *Journal of Hydraulic Engineering* 143(5): 04017001.
- 27) Noori, R., Deng, Z., Kiaghadi, A., Kachoosangi, F. T. 2015. How reliable are ANN, ANFIS, and SVM techniques for predicting longitudinal dispersion coefficient in natural rivers? *Journal of Hydraulic Engineering*. 142(1): 04015039.
- 28) Noori, R., Karbassi, A. R., Ashrafi, K., Ardestani, M., Mehrdadi, N. 2013. Development and application of reduced-order neural network model based on proper orthogonal decomposition for BOD5 monitoring: Active and online prediction. *Environmental progress & sustainable energy* 32(1): 120-127.
- 29) Noori, R., Khakpour, A., Omidvar, B., Farokhnia, A. 2010. Comparison of ANN and principal component analysis-multivariate linear regression models for predicting the river flow based on developed discrepancy ratio statistic. *Expert Systems with Applications* 37(8): 5856-5862.
- 30) Noori, R., Karbassi, A., Farokhnia, A., & Dehghani, M. 2009. Predicting the longitudinal dispersion coefficient using support vector machine and adaptive neuro-fuzzy inference system techniques. *Environmental Engineering Science*. 26(10): 1503-1510.
- 31) Noori, R., Farokhnia, A., Morid, S., and Riahi Madvar, H. 2009. Effect of input variables preprocessing in artificial neural network on monthly flow prediction by PCA and wavelet transformation. *Journal of Water and Wastewater*. 69: 13-22. (In Persian)
- 32) Noori, R., Abdoli, M. A., Ameri, A., and Jalili-Ghazizade, M. 2008. Prediction of municipal solid waste generation with combination of support vector machine and principal component analysis: A case study

Water resources management. 27(9): 3307-3318.

43) Zeng, Y., and Wenxin, H. 2014. Estimation of longitudinal dispersion coefficient in rivers. Journal of Hydro-environment Research 8(1): 2-8.

