

بهینه سازی شبکه عصبی ELM در مسئله پیش بینی

جلال رضایی نور^{1*}، منصوره یاری ایلی²، اسماعیل هداوندی³، محمدحسین روزبهانی⁴

¹ دانشیار، گروه مهندسی صنایع، دانشگاه قم، قم، ایران (عهده دار مکاتبات)

² دانشجوی دکتری، گروه مهندسی کامپیوتر و فناوری اطلاعات، دانشگاه قم، قم، ایران

³ استادیار، گروه مهندسی صنایع و کامپیوتر، دانشگاه صنعتی بیرجند، بیرجند، ایران

⁴ دانش آموخته دکتری، گروه مهندسی مکانیک، دانشگاه تربیت مدرس، تهران، ایران

تاریخ دریافت: بهمن 1397، اصلاحیه: اردیبهشت 1398، پذیرش: مرداد 1398

چکیده

برای محیط همیشه در حال تغییر این دوران و تغییر در شیوه تعاملات شرکت‌ها با تامین‌کنندگان و مشتریان و پیچیدگی بازارها، کاهش دوره عمر محصولات و اهمیت یافتن زمان پاسخگویی به مشتریان، پیش‌بینی تقاضای محصول عامل حیاتی برای رقابت‌پذیری سازمان‌ها می‌باشد. با پیش‌بینی دقیق الگوهای تقاضا، دولت و تولیدکنندگان می‌توانند براساس میزان و جهت جریان آتی تولید، برنامه ریزی‌ها را سازماندهی کنند و زیرساخت بهتری در رشد صنعت مهیا سازند. با هدف ارائه یک ابزار دقیق پیش‌بینی در صنعت فولاد، این مطالعه، به بهینه سازی شبکه عصبی ELM به کمک الگوریتم ژنتیک می‌پردازد. که در آن پارامترهای شبکه، از قبیل تعداد و توابع فعال‌سازی نورون‌ها در لایه مخفی، وزن‌های اتصالی بین ورودی‌ها و نورون‌های لایه مخفی، بایاس نورون‌های لایه مخفی و σ (پارامتر تنظیم‌سازی)، به کمک الگوریتم ژنتیک تعیین می‌شود. برای پیش‌بینی تقاضای فولاد خام کشور، داده‌های مرتبط با تولید و مصرف فولاد خام و محصولات فولادی کشور، بصورت ماهیانه و در بازه زمانی دی ماه 88 تا مرداد ماه 92 (جمعاً 44 نمونه و 22 مشخصه) جمع‌آوری گردید. داده‌های دی ماه 88 تا بهمن 91 در آموزش شبکه و داده‌های مربوط به اسفند 91 تا مرداد 92 در آزمون شبکه استفاده شد. برای نشان دادن کارایی مدل پیش‌بینی کننده، مقایسه‌ی عملکرد از لحاظ دقت پیش‌بینی و سرعت یادگیری بین الگوریتم ELM بهینه شده با الگوریتم ELM و سایر شبکه‌های عصبی صورت گرفته است. معیارهای پیش‌بینی نشان‌دهنده‌ی عملکرد خوب ELM بهینه شده نسبت به سایر شبکه‌های عصبی می‌باشد. براساس آزمون‌های آماری و خطاهای RMSE و MAPE نتایج نشان می‌دهد که دقت ELM به مراتب بهتر از سایر روش‌های شبکه عصبی است. به علاوه، مدل ELM دو بار سریعتر از شبکه‌های عصبی کلاسیک است. براساس یافته‌ها می‌توان به یقین گفت که بین الگوریتم‌های مورد بررسی، ELM ابزار دقیق‌تر و قوی‌تری در مسئله تقاضای فولاد است. البته نوع بهینه یافته شبکه عصبی ELM دقت بهتری در مدل‌سازی تابع تقاضا داشته است، ولی روش ELM از نظر زمانی بهینه‌تر بوده است.

کلمات کلیدی: شبکه‌های عصبی مصنوعی، الگوریتم ژنتیک، پیش‌بینی، تقاضای فولاد خام

1- مقدمه

چگونگی پیش‌بینی بازار عرضه و تقاضای محصولات یک موضوع مهم و پرطرفدار بوده و توجه محققان زیادی را به خود جلب کرده است [3,7]. به‌طور کلی برای مدل‌سازی عرضه و تقاضا دو مدل کلی در مطالعات، مورد بررسی محققان قرار گرفته است: مدل‌های آماری و مدل‌های یادگیری ماشین [13]. با توجه به مطالعات گذشته تحلیل‌های رگرسیون آماری به عنوان پرطرفدارترین روش در پیش‌بینی عرضه و تقاضای محصولات مورد توجه قرار گرفته است. اما مدل‌های یادگیری ماشین دارای مزیت‌هایی نظیر محاسبه سریع‌تر و دقیق‌تر، هزینه پایین و سهولت در پیاده‌سازی می‌باشد. الگوریتم یادگیریم‌اشین که به‌طور مکرر در مدل‌سازی توابع عرضه و تقاضای مختلف، مورد استفاده قرار گرفته شبکه‌های عصبی می‌باشد [4,9] که ابزاری قدرتمند برای مدل‌سازی ورودی و یافتن رابطه بین ورودی و خروجی است.

در تصمیم‌گیری‌های آینده محور درون سازمانی، برنامه‌ریزی و کنترل نقش بسیار مهم و حیاتی ایفا می‌کند. پیش‌بینی تغییرات الگوها و روندهای آتی در ساختار عرضه و تقاضای محصولات، پایه و اساس برنامه‌ریزی است. تصمیم‌گیری‌های راهبردی مدیران ارشد و برنامه‌ریزی-های کلان سازمان‌ها بر مبنای اطلاعات به‌دست آمده از پیش‌بینی‌ها انجام می‌شود. در صورتی که اطلاعات پیش‌بینی عرضه و تقاضا از دقت و صحت کافی برخوردار نباشد هزینه‌هایی نظیر مواجه شدن با موجودی مازاد غیرقابل استفاده، هزینه‌های نگهداری موجودی و... بر سازمان تحمیل خواهد شد؛ در نتیجه سازمان قابلیت رقابتی‌اش را از دست خواهد داد. به‌عنوان نتیجه می‌توان گفت بهبود عملکرد سازمانی، بستگی به دقت پیش‌بینی عرضه و تقاضا دارد. به همین دلیل در سال‌های اخیر،

*j.rezaee@qom.ac.ir

نام ELM¹، اخیراً توسط هوآنگ ارائه شد که سرعت خوبی در یادگیری دارد [21] و در آن وزن‌های ورودی (که لایه ورودی را به لایه خروجی متصل می‌کنند) و بایاس‌های مخفی به صورت تصادفی انتخاب می‌شوند. ELM نه تنها نسبت به الگوریتم‌های یادگیری سنتی براساس گرادایان، خیلی سریع‌تر آموزش می‌بیند بلکه از بسیاری از پیچیدگی‌هایی که آن الگوریتم‌ها دارند (نظیر معیار توقف، نرخ یادگیری، epochها) اجتناب می‌کند. شبکه عصبی ELM به خاطر مزایای چشمگیر مثل سادگی، عملکرد تعمیم و سرعت اجرایی بالا توجه محققان زیادی را در زمینه‌های تحقیقاتی گوناگون به خود معطوف کرده است. به غیر از کلاس‌بندی و رگرسیون [22] اخیراً برای خوشه‌بندی و انتخاب مشخصه نیز استفاده شده است [22] ELM به طرز موفقیت‌آمیزی در زمینه‌های بیوانفورماتیک و بهداشت و درمان الکترونیک، [31] پردازش سیگنال و تصویر [۱۱، ۲۸]، تشخیص حرکت در انسان، [27] انتخاب مشخصه، [12] پیش‌بینی سری زمانی، [17] آموزش و پیش‌بینی به صورت بلادرنگ، [26] کنترل و رباتیک [36] و... به کار گرفته شده است. ELM در مقایسه با ماشین بردار پشتیبان هم کارآمدتر است. مطالعات نشان داده که توانایی تعمیم‌یابی این شبکه‌ها نسبت به ماشین بردار پشتیبان و انواع آن بهتر است. [27, 28]

لذا با توجه به خلا موجودی که در زمینه‌ی به‌کارگیری روش‌های پیشرفته در پیش‌بینی صنعت فولاد در کشور حس می‌شود این پژوهش سعی بر آن دارد که با استفاده از فنون نوین داده کاوی، به مدلسازی ماهیانه‌ی تابع تقاضای فولاد بپردازد و راه‌حلی کارآمد برای این مسئله ارائه دهد. این تحقیق سه هدف را دنبال می‌کند: هدف اول ارزیابی قدرت و دقت مدل ELM در مسئله پیش‌بینی ماهیانه تقاضای فولاد به کمک شاخص‌هایی جدید به عنوان متغیرهای ورودی است (که این شاخص‌ها تا بحال در مطالعات پیشین مورد بررسی قرار نگرفته‌اند). هدف دوم بهینه سازی یادگیری شبکه ELM با الگوریتم ژنتیک است. در این روش، تعداد و توابع فعالسازی نورون‌ها در لایه مخفی، وزن‌های اتصالی بین ورودی‌ها و نورون‌های لایه مخفی و بایاس نورون‌های لایه مخفی با فاکتور تنظیم سازی Tikhonov بهبود می‌یابند. هدف سوم تحلیل آماری نتایج و مقایسه خروجی‌های ELM با خروجی‌های سایر شبکه‌های عصبی است.

2- تئوری اولیه ELM

همانطور که می‌دانیم سرعت شبکه‌های عصبی پس انتشار² خیلی کمتر از حد مورد نیاز است و در دهه‌های اخیر این نقطه ضعف، به یک گلوگاه در کاربردهای عملی تبدیل شده است. دو دلیل عمده برای این سرعت کم وجود دارد: اول الگوریتم‌های یادگیری کند، بر پایه گرادایان، جهت آموزش شبکه عصبی استفاده شده است و دوم در این نوع الگوریتم‌های یادگیری همه پارامترها باید مکرراً تنظیم شوند. [23] به

در رابطه با موضوع خاص موجود، یعنی تولید و مصرف فولاد، اطلاعات متفاوتی از مراجع مختلف عنوان می‌شود که به نظر می‌رسد از شفافیت و نیز تحلیل‌های مناسب برخوردار نیست. دلیل این تناقض، عدم وجود یک مدل پیش‌بینی دقیق و مناسب است که موجب ناهمگونی نتایج می‌شود. بسیاری از مقالاتی که در زمینه صنعت فولاد منتشر شده‌اند از تکنیک‌های سری‌های زمانی و اقتصادسنجی استفاده کرده‌اند. به عنوان مثال در ایران در دو مطالعه‌ی آذربایجانی (1380) و گودرزی (1386)، تخمین تابع تقاضای فولاد و پیش‌بینی آن با روش‌های مذکور انجام شده است [8، 1]. علاوه بر آن برخی از مطالعات خارجی صورت گرفته در این حوزه عبارت است از مطالعه‌ی زیانگین (2013) که به پیش‌بینی میزان مصرف فولاد تا سال 2050 در چین پرداخته است [35]. در مطالعه کرامپتن (2000) پیش‌بینی مصرف فولاد در ژاپن با تکنیک شدت استفاده به صورت مجزا برای شش صنعت مصرف‌کننده فولاد (تولید ماشین‌آلات و ابزارآلات حمل و نقل، مسکن و...) تا سال 2005 صورت گرفته است [29]. علاوه بر آن رابرت (1988) با استفاده از صادرات و درآمد ملی در آمریکا، فیندلی (1985) با استفاده از متغیرهای درآمد ملی و صادرات در ژاپن، با استفاده از روش مشابه (شدت استفاده) به پیش‌بینی تابع تقاضای فولاد پرداخته‌اند [32، 34]. همان‌طور که مشاهده می‌شود همه مطالعات فوق از روش‌های اقتصادسنجی و سری‌زمانی بهره‌جسته‌اند. تکنیک‌های سری زمانی از یک تابع عملکردی خطی استفاده می‌کنند. در بسیاری از داده‌های تقاضا، رابطه بین متغیر هدف و متغیرهای مستقل، غیر خطی و پیچیده است. بسیار مشکل یا حتی غیرممکن است که بتوانیم مدل‌های ریاضی برای توصیف این روابط تعریف کنیم [33]. این مدل‌ها همچنین به داده‌های تاریخی بسیار بزرگی برای رسیدن به نتایج دقیق نیاز دارند. از طرفی داده‌های ناکامل و یا گم‌شده یکی از مشکلات معمول در بسیاری از مسائل داده کاوی و یادگیری ماشین است که می‌بایست عملاً به آن‌ها توجه نمود. دلایل متفاوتی برای ناکامل بودن مجموعه داده‌ها وجود دارد. برای مثال در بخش صنعت عدم گزارش داده‌ها توسط منابع مربوطه و اختلال در عملکرد دستگاه‌ها باعث نقص در مجموعه داده می‌شود. بنابراین، مدل‌های پیش‌بینی امروزه نیاز است که بتوانند تحت شرایط داده‌های ناکامل و یا داده‌های با ابعاد کم نیز کارا باشند [15].

گرچه هنوز بیش از 50 سال از تولد روش‌های محاسباتی مبتنی بر شبکه‌های عصبی مصنوعی نمی‌گذرد، لکن این شبکه‌ها به دلیل ویژگی‌هایی همچون پردازش موازی، هوشمندی و انعطاف‌پذیری در حل مسائل غیرخطی جایگاه چشم‌گیری در حل مسائل پیچیده از قبیل شناخت الگو، خوشه‌بندی، تخمین و شناسایی، مدل‌سازی و پیش‌بینی برای خود باز کرده‌اند. در موضوع مورد بحث این پژوهش شبکه‌های عصبی در مقایسه با روش‌های کلاسیک همچون هموارسازی نمایی، بهتر می‌توانند رفتار غیرخطی، نوسانی و حتی آشوبی تقاضای متلاطم را مدل‌سازی و پیش‌بینی نمایند [2].

یک الگوریتم یادگیری شبکه عصبی پس انتشار با یک لایه مخفی به

¹ Extreme Learning Machine

² FFNN

$$y = g\left(b_o + \sum_{j=1}^h w_{jo} v_j\right) \quad (1)$$

$$v_j = f_j\left(b_i + \sum_{i=1}^n a_{ij} x_i\right)$$

n : تعداد متغیرهای ورودی و تعداد نورون های لایه مخفی است. v_j خروجی نورون های لایه مخفی است و $j = 1, \dots, h$ متغیرهای ورودی هستند. a_{ij} وزن اتصالات بین متغیر ورودی i و نورون j در لایه مخفی است. w_{jo} وزن اتصالات بین نورون j در لایه مخفی و نورون خروجی است. b_i بایاس نورون در لایه مخفی و b_o بایاس نورون خروجی است. f_j و g تابع فعال سازی نورون j در لایه مخفی و تابع فعال سازی نورون خروجی است. x_i یک متغیر باینری است که در گزینش متغیرهای ورودی در خلال طراحی SLFN به کار می رود. با استفاده از این متغیر باینری هر متغیر ورودی را می توان مورد بررسی قرار داد (که آیا در ساختار شبکه باشد یا خیر). [18]

ELM، شبکه SLFN را در دو مرحله آموزش می دهد: 1- نگاشت مشخصه تصادفی و 2- حل پارامترهای خطی. در مرحله اول به طور تصادفی لایه پنهان را مقداردهی اولیه می کند تا داده ورودی را به فضای مشخصه نگاشت کند. این کار توسط بعضی توابع نگاشت غیر خطی انجام می شود [21]. (شکل 2) مرحله نگاشت تصادفی مشخصه، ELM را از بسیاری از الگوریتم های یادگیری موجود مثل SVM⁴ (که از توابع kernel برای نگاشت مشخصه استفاده می کنند) یا Deep Neural Network (که از RBM⁵ یا AE⁶ برای آموزش مشخصه استفاده می کند) مجزا می کند. توابع نگاشت غیرخطی در ELM می تواند هر تابع پیوسته غیرخطی باشد. فهرست برخی توابعی که اغلب استفاده می شوند در جدول 1 آمده است [23].

شکل 2 ساختار یک شبکه عصبی ELM را نشان می دهد. تابع خروجی ELM جهت تعمیم SLFN به صورت زیر است:

$$f_L(x) = \sum_{i=1}^L \beta_i h_i(x) = h(x)\beta \quad (2)$$

که در آن $\beta = [\beta_1, \dots, \beta_L]^T$ بردار وزن خروجی بین L گره لایه پنهان و m گره خروجی است $h(x) = [h_1(x), \dots, h_L(x)]$ ، $(m \geq 1)$ نگاشت مشخصه غیر خطی ماشین یادگیری فرین است. عمل h_i به صورت زیر توصیف می شود:

$$h_i(x) = G(a_i, b_i, x), \quad a_i \in \mathbb{R}^d, \quad b_i \in \mathbb{R} \quad (3)$$

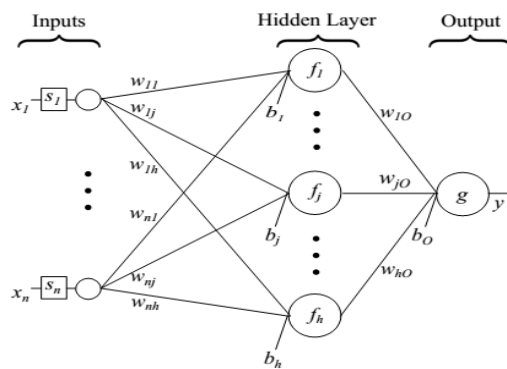
که $G(a, b, x)$ تابع پیوسته غیرخطی تکه ای⁷ است. در شبکه عصبی ELM پارامترهای گره مخفی یعنی a, b به صورت تصادفی (مستقل از داده آموزشی) و براساس توابع توزیع احتمالی تولید می شوند. این ماهیت منجر به کارایی بیشتر در مقایسه با شبکه های

همین دلیل در کاربردهای صنعتی و عملی، مدل های خطی اغلب به خاطر سرعت بالاتر در یادگیری، به شبکه های عصبی پس انتشار ترجیح داده می شوند. به منظور غلبه بر ساختار کند شبکه های پس انتشار، یک شبکه عصبی جدید به نام ELM ارائه شد. [21] این شبکه عصبی، تعمیم یافته شبکه های پس انتشار تک لایه است اما تفاوت اصلی آن با سایر شبکه ها ناشی از این است که برخلاف روش های یادگیری معمولی که قبل از تولید پارامترهای نورون های مخفی باید داده آموزشی را ببینند؛ در شبکه عصبی ELM وزن های ورودی (وزن اتصالات بین متغیرهای ورودی و نورون های لایه پنهان) و بایاس نورون های لایه مخفی به طور تصادفی انتخاب می شوند و ELM می تواند قبل از آن که داده های یادگیری را ببیند این پارامترها را تولید کند. در ادامه به توضیح اساس تئوری ELM پرداخته می شود. همانطور که

جدول (1): فهرست توابع نگاشت غیرخطی در ELM

Sigmoid Function	$G(a, b, x) = \frac{1}{1 + \exp(-(a \cdot x + b))}$
Hyperbolic tangent Function	$G(a, b, x) = \frac{1 - \exp(-(a \cdot x + b))}{1 + \exp(-(a \cdot x + b))}$
Gaussian Function	$G(a, b, x) = \exp(-b \ x - a\)$
Multiquadric Function	$G(a, b, x) = (\ x - a\ + b^2)^{1/2}$
Hard Limit Function	$G(a, b, x) = \begin{cases} 1, & \text{if } a \cdot x + b \leq 0 \\ 0, & \text{Otherwise} \end{cases}$
Cosin Function/ Fourier basis	$G(a, b, x) = \cos(a \cdot x + b)$

گفته شد شبکه عصبی ELM تعمیم یافته SLFN³ است. شکل 1 ساختار SLFN را نشان می دهد که در قالب ریاضی به صورت زیر فرموله می شود: [21]



شکل (1): ساختار SLFN

³ Single Hidden Layer Feed Forward Neural Network

⁴ Support Vector Machine

⁵ Restricted Boltzmann Machine

⁶ Auto-Encoders/Auto-Decoders

⁷ Nonlinear Piecewise Continuous Function

مرسوم (که براساس تنظیم سازی هستند)؛ گرایش به تعداد نورون‌های زیاد دارد. هم‌چنین این شبکه‌ها در حضور مقادیر ورودی نامرتب و هم-بسته با مشکلات بیش برآزش مواجه می‌شوند.

3- بهینه سازی ELM با الگوریتم ژنتیک

در این بخش، به ارائه مدل پیشنهادی این مطالعه خواهیم پرداخت.

1-3 الگوریتم ژنتیک

الگوریتم های تکاملی¹⁰ روش‌های جست‌وجو و بهینه‌سازی هستند که براساس انتقال خصوصیات موروثی توسط ژن‌ها و طبق نظریه داروین عمل می‌کنند و سعی در یافتن راه حل بهینه‌ای برای حل مسائل پیچیده دارند. اساس کار الگوریتم ژنتیک، جستجوی تصادفی است. با این وجود الگوریتم ژنتیک می‌تواند در اکثر موارد به نزدیک جواب بهینه همگرا شود. الگوریتم ژنتیک به جای شروع از یک نقطه در مجموعه جواب‌ها، از چندین نقطه به طور همزمان به جستجوی پاسخ می‌پردازد که این باعث جلوگیری از افتادن الگوریتم در مینیمم های محلی¹¹ می‌شود. همچنین با خود تابعی که باید مینیمم شود روبرو می‌شود و نیازی به مشتقات آن ندارد، لذا شامل پیچیدگی‌های محاسباتی، یا محاسبات اضافی نیست. طرح کلی الگوریتم ژنتیک به شرح زیر می‌باشد:

(الف) کدبندی کروموزوم‌ها و تولید جمعیت اولیه

الگوریتم ژنتیک به جای اینکه روی متغیرهای مسأله کار کند، با شکل کد شده آن‌ها سروکار دارد. برای کمینه کردن هزینه تابعی که شامل متغیرهای حقیقی و صحیح و باینری، همه متغیرها به متغیرهای پیوسته و حقیقی به وسیله تکنیک‌های مختلف کدگذاری (از جمله باینری، جایگشتی، درخت و...) نگاشت می‌شود و هر کروموزوم به صورت یک عدد حقیقی در بازه [0,1] کدگذاری می‌شود. جمعیت اولیه، یا راه‌حل‌های مناسب مسئله، به صورت تصادفی و با توزیع یکنواخت انتخاب می‌شود. تعداد اعضای جمعیت متناسب با طول کروموزوم‌ها است. اگر طول کروموزوم یا تعداد متغیرهای مجهول مسأله بیشتر شود باید جمعیت اولیه نیز بزرگتر باشد. هر کروموزوم نشان دهنده یک جواب ممکن برای بهینه سازی مسأله است

$$P = \{P_1, \dots, P_m\} \quad (8)$$

که

$$P_k = [P_{k1}, P_{k2}, \dots, P_{kq}]^T \quad (9)$$

P_{ik} متغیر i در کروموزوم k است .

m, q ($i = 1, \dots, q, k = 1, \dots, m, 0 \leq P_{ik} \leq 1$) به ترتیب اندازه

جمعیت و تعداد متغیرها هستند.

(ب) ارزیابی¹²

دو کروموزوم (والدین) با توجه به برآزندگی آنها از میان جمعیت انتخاب می‌شوند (هر چه برآزندگی بیشتر باشد شانس انتخاب بیشتر است) تابع برآزندگی را از اعمال تبدیل مناسب بر روی تابع هدف یعنی تابعی که

عصبی سنتی می‌شود. در مرحله دوم از یادگیری، β (وزن های بین لایه پنهان و لایه خروجی) با به حداقل رساندن خطای مربعاتی به دست می‌آید.⁸

$$\beta \in R^{L \times m} \text{ که } \min_{\beta} \|H\beta - T\|^2 \quad (4)$$

که در آن H ماتریس خروجی لایه پنهان است :

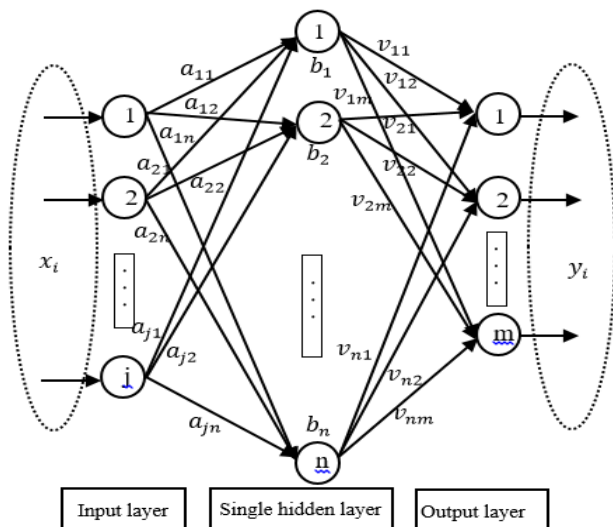
$$H = \begin{bmatrix} h_1(x_1) \\ \vdots \\ h_l(x_1) \\ \vdots \\ h_1(x_N) \\ \vdots \\ h_l(x_N) \end{bmatrix} = \begin{bmatrix} h_{11}(x_1) & \dots & h_{1l}(x_1) \\ \vdots & \ddots & \vdots \\ h_{11}(x_N) & \dots & h_{1l}(x_N) \end{bmatrix} \quad (5)$$

و T ماتریس هدف داده آموزشی است:

$$T = \begin{bmatrix} t_1^T \\ \vdots \\ t_N^T \end{bmatrix} = \begin{bmatrix} t_{11} & \dots & t_{1m} \\ \vdots & \ddots & \vdots \\ t_{N1} & \dots & t_{Nm} \end{bmatrix} \quad (6)$$

جواب بهینه برای β به قرار زیر است⁹:

$$\beta = H^{-1}T \quad (7)$$



شکل (2): ساختار کلی ELM

از نقطه نظر یادگیری و بهره‌وری، ماشین یادگیری فرین سه هدف را دنبال می‌کند: حداقل دخالتهای انسانی، دقت بالای یادگیری و سرعت بالای یادگیری. اما از آن جا که تابع فعال‌سازی نورون‌های خروجی خطی است (چون در شبکه عصبی ELM وزن‌های خروجی براساس وزن‌های ورودی تصادفی و بایاس گره‌های مخفی محاسبه می‌شوند) ممکن است مجموعه‌ای از وزن‌های ورودی غیربهینه یا گره‌های مخفی غیرضروری وجود داشته باشد. به علاوه ELM نسبت به الگوریتم‌های یادگیری

⁸ به نرم Frobenius اشاره دارد
⁹ H^{-1} معکوس ماتریس H

¹⁰ Evaluation Algorithms

¹¹ Local Minimum

¹² Evaluation

کنید: [10,18]

$$\begin{aligned} \text{Parent1} &= p_{1_1} \quad p_{1_2} \quad p_{1_3} \quad p_{1_4} \\ \text{Parent2} &= p_{2_1} \quad p_{2_2} \quad p_{2_3} \quad p_{2_4} \\ \text{Mask} &= 1 \quad 0 \quad 0 \quad 1 \\ \text{Offspring1} &= p_{1_1} \quad p_{2_2} \quad p_{2_3} \quad p_{1_4} \\ \text{Offspring2} &= p_{2_1} \quad p_{1_2} \quad p_{1_3} \quad p_{2_4} \end{aligned}$$

با توجه به احتمال ترکیب شدن، والدین برای تشکیل فرزندان¹⁴ جدید با هم ترکیب می شوند. ترکیب مهم ترین عملگر بوده و فرآیندی است که در آن نسل قدیمی کروموزومها با یکدیگر مخلوط می-شوند تا نسل تازه‌ای از کروموزومها بوجود بیاید. جفت‌هایی که در قسمت انتخاب به عنوان والد در نظر گرفته شدند در این قسمت ژن‌هایشان را با هم مبادله می‌کنند و اعضای جدید را بوجود می‌آورند. ترکیب در الگوریتم ژنتیک باعث از بین رفتن پراکندگی یا تنوع ژنتیکی جمعیت می‌شود زیرا اجازه می‌دهد ژن‌های خوب یکدیگر را بیابند.

(د) عملیات جهش¹⁵

برای اینکه از گیر افتادن الگوریتم در بهینه‌های محلی اجتناب شود و نیز تنوع جمعیت و گستردگی فضای جست و جو حفظ شود، عملگر جهش به کار می‌رود. اگر نرخ جهش خیلی زیاد باشد موجب کندی یا واگرایی الگوریتم می‌شود. هم‌چنین نباید آن را صفر در نظر گرفت. برای هر فرزند یک عدد تصادفی r تولید می‌شود اگر $r \leq \eta_{lk}$ احتمال جهش است. این فرزند جهش یافته است. جهش یک عملگر 2 مرحله‌ای است؛ ابتدا یک عنصر تصادفی از فرد با یک مقدار تصادفی در بازه [0,1] جایگزین می‌شود. اگر عنصر دوم که برای جایگزینی انتخاب شده بود کروموزوم جهش یافته باشد به صورت زیر توصیف می‌شود:

$$p_{k_2}^i = [p_{k_2,1}^i, p_{k_2,2}^i, p_{k_2,3}^i, p_{k_2,4}^i]^T \quad (14)$$

که $p_{k_2}^i$ یک عدد تصادفی در بازه [0,1] است. در گام دوم فاکتور تطبیق تصادفی به کروموزوم اضافه می‌شود. فاکتور تطبیق حاصل ضرب هر المان 1 با کروموزوم جهش یافته قبلی $p_{k_2}^i$ با یک عدد تصادفی $-1 \leq \beta_{lk} \leq 1$ و در نهایت ضرب نتیجه کروموزوم با فاکتور جهش $0 \leq \eta_{lk} \leq 1$ است. بنابراین داریم:

$$p_{k_2}^i = \eta_{lk} [\beta_{lk,1} p_{k_2,1}^i, \beta_{lk,2} p_{k_2,2}^i, \beta_{lk,3} p_{k_2,3}^i, \beta_{lk,4} p_{k_2,4}^i]^T \quad (15)$$

دست آخر کروموزوم جهش یافته به صورت زیر است:

$$p_{k_2}^i = \text{rem}(p_{k_2}^i + p_{k_2}^i) \quad (16)$$

باقی مانده هر متغیر پس از تقسیم شدن بر یک است. با توجه به احتمال جهش¹⁶ فرزندان در هر لوکاس (موقعیت در کروموزوم)¹⁷ مورد

قرار است بهینه شود به دست می‌آورند. این تابع هر رشته را با یک مقدار عددی ارزیابی می‌کند که کیفیت آن را مشخص می‌نماید. هر چه کیفیت رشته جواب بالاتر باشد مقدار برازندگی جواب بیشتر است و احتمال مشارکت برای تولید نسل بعدی نیز افزایش خواهد یافت. هر کروموزوم می‌تواند با استفاد از تابع *fitness* که برای حل مسئله تخصیص داده شده ارزیابی گردد. از آن جا که همه متغیرها به مقادیر پیوسته بین [0,1] نگاشت شده اند، قبل از به دست آوردن *fitness* هر فرد این مقادیر باید به مقادیر واقعی خود تبدیل شوند. اگر مقدار اصلی متغیر l ام ($l = 1, \dots, v$) در فرد k ($k = 1, \dots, m$) حقیقی باشد داریم:

$$z_{k_l} = (z_l^{\max} - z_l^{\min}) p_{k_l} + z_l^{\min} \quad (10)$$

که z_l^{\max} و z_l^{\min} مقدار درست بازه های متغیرها را نشان می‌دهد. ($z_l^{\min} \leq z_{k_l} \leq z_l^{\max}$). اگر یک مقدار صحیح باشد داریم:

$$z_{k_l} = \text{rounddown}((z_l^{\max} - z_l^{\min} + 1) p_{k_l}) + z_l^{\min} \quad (11)$$

rounddown تابعی است که به بزرگترین عدد صحیح که کوچکتر مساوی آرگومان است گرد می‌کند. اگر مقدار واقعی متغیر باینری باشد داریم:

$$z_{k_l} = \text{round}(p_{k_l}) \quad (12)$$

round تابعی است که به نزدیکترین عدد صحیح گرد می‌کند. پس از تبدیل متغیرها *fitness* هر فرد می‌تواند به دست آید. تابع *fitness* که یک کروموزوم را در جمعیت ارزیابی میکند به صورت زیر نوشته می‌شود:

$$\text{fitness} = \psi(p_k) \in \mathcal{R} \quad (13)$$

تابع ارزیابی $\psi(\cdot)$ برای مسئله‌ای که قرار است حل شود تخصیص یافته است. براساس مقادیر *fitness* مجموعه‌ای از افرادی که برای نسل بعد درامان مانده اند انتخاب می‌شوند و بقیه حذف می‌شوند. بهترین افراد، باقی مانده و جای کروموزوم‌های حذف شده را نسل جدید می‌گیرند. برای گزینش پدر و مادر از گزینش Tournament استفاده شده است. [32]

(ج) عملیات ادغام¹³ دو کروموزوم

از عملگر ترکیب *uniform* استفاده کردیم چرا که نسبت به عملگرهای دیگر ترکیب، فضای اکتشافی وسیع تری ارائه می‌دهد. در این عملگر ابتدا یک ماسک باینری تصادفی که طول آن مساوی طول فرد است تولید می‌شود. سپس هر فرزند مقادیری را از والد اول یا دومش می‌گیرد. این موضوع بسته به این دارد که بیت ماسک صفر است یا یک. اگر بیت ماسک یک باشد فرزند 2/1 مقادیر را از والد 2/1 می‌گیرد. اگر بیت ماسک صفر باشد فرزند، مقادیر را از والد 1/2 می‌گیرد. به مثال زیر توجه

¹³ Cross over

¹⁴ Offspring

¹⁵ Mutation

¹⁶ Probability Mutation

¹⁷ Lockus

در فرآیند بهینه سازی، هر کروموزوم $P_k, k = \{1, \dots, m\}$ در جمعیت P از موارد زیر تشکیل شده است:

$$P_k = [\omega_{11}^k, \dots, \omega_{1n}^k, h_1^k, \dots, h_n^k, s_{11}^k, \dots, s_{1n}^k, \alpha^k] \quad (22)$$

$$k = 1, \dots, m$$

تابع فعال سازی f_j (مربوط به هر نورون j در لایه مخفی) به نحو زیر تعریف می شود:

$$s_j^A = \begin{cases} 0 & \text{if } s_j^A = 0 \\ 1/(1 + \exp(-v)) & \text{if } s_j^A = 1 \\ v & \text{if } s_j^A = 2 \end{cases} \quad (23)$$

یک عدد صحیح است که استفاده از آن، امکان تنظیم تعداد نورون ها ($s_j^A = 0$ نورون لحاظ نشده است) و تابع فعال سازی هر نورون را فراهم می کند. در این پژوهش فقط دو نوع تابع فعال سازی (سیگموئیدی و خطی) استفاده شده است. همان طور که در بخش قبلی گفته شد همه متغیرها به متغیرهای حقیقی در بازه $[0, 1]$ نگاشت می شوند پس قبل از تخمین *fitness* هر فرد، متغیرها باید به مقادیر حقیقی خود تبدیل شوند. متغیرهای $s_{j,i} = 1, \dots, n$ باینری هستند و با استفاده از معادله 12 تبدیل می شوند. متغیرهای $h_{j,i} = 1, \dots, n$ صحیح بوده و با فرمول 11 تبدیل می شوند. (بازه های بالایی و پایینی 0 و 2) وزن های ورودی w_{ij} و بایاس b_j (بازه بالایی و پایینی 1 و -1) و در نهایت α (بازه بالا و پایینی 0 و 100) با معادله 13 حاصل می شود.

4- بحث

1-4 مجموعه داده های مورد استفاده در این پژوهش

در اولین مرحله بایستی لیستی از متغیرهای تاثیرگذار بر عرضه فولاد در کشور شناسایی شود که با بررسی مقالاتی که در این زمینه وجود دارد و مشورت با خبرگان مواردی به عنوان عوامل اثرگذار شناسایی گردید. در مرحله دوم گردآوری داده ها از منابع مختلف داخلی (از جمله مرکز آمارهای اقتصادی وزارت صنایع، بانک مرکزی، مرکز آمار ایران و ...) و همچنین منابع خارجی¹⁹ مد نظر قرار گرفت. در این مرحله داده های ماهانه عرضه، تقاضا، واردات و صادرات فولاد خام، محصولات فولادی، ورق گرم و میلگرد که مورد نیاز بود جمع آوری گردید. داده های فوق بصورت ماهیانه و در بازه زمانی دی ماه 88 تا مرداد ماه 92 و جمعاً 44 داده موجود بود. با توجه به اینکه داده های سالانه تنها در 10 سال اخیر وجود داشت و عملاً تعداد آنها برای انجام یک مدل سازی مناسب بسیار کم بود، مدل سازی ماهانه تابع فولاد خام مد نظر قرار گرفت. بنابراین آمار مشخصه های دیگر در این بازه زمانی بایستی جمع آوری می گردید. اطلاعات مربوط به قیمت های فولاد خام و قراضه به تومان در کشور

¹⁹ منابع متعدد خارجی از جمله <http://www.steelonthenet.com/semi-finished-prices.php>

جهش قرار می گیرند. در الگوریتم ژنتیک بعد از این که یک عضو در جمعیت جدید بوجود آمد هر ژن آن با احتمال جهش، جهش می یابد. در نهایت بعد از این که الگوریتم بهترین جواب را برای مسأله ارائه کرد لازم است عکس عمل رمزگذاری (رمزگشایی) روی جواب ها اعمال شود تا بتوانیم نسخه واقعی جواب را به وضوح در دست داشته باشیم. [10,19]

2-3 مدل پیشنهادی

خواص جست و جو و بهینه سازی الگوریتم ژنتیک آن ها را برای طراحی ساختار و وزن های مدل مناسب کرده است؛ به خصوص در شبکه های عصبی پس انتشار. در این پژوهش پارامترها و ساختار ELM به کمک الگوریتم ژنتیک توصیف شده در بخش قبل بهینه می شود. همانطور که قبلاً گفتیم ELM نسبت به الگوریتم های برپایه تنظیم سازی دستی، گرایش به تولید گره های مخفی بیشتری دارد. به علاوه حضور متغیرهای بی ربط در مجموعه داده آموزشی باعث کاهش عملکرد می شود. برای غلبه بر این مشکلات در روش پیشنهادی تعداد و توابع فعال سازی نورون ها در لایه مخفی، وزن های اتصالی بین ورودی ها و نورون های لایه مخفی، بایاس نورون های لایه مخفی و α (پارامتر تنظیم سازی) به کمک الگوریتم ژنتیک تعیین می شوند. که در ادامه تشریح می شود.

در روش حداقل مربعاتی، هدف یافتن بهترین وزن خروجی به صورت زیر است:

$$\min(\|y - y_d\|_2) \quad (17)$$

مقاومت راه حل حداقل مربعاتی در مقابل نویز با روش تنظیم سازی Tikhonov بیشتر می شود. این مسئله با استفاده از جمله تنظیم سازی Tikhonov تبدیل به مسئله زیر می شود:

$$\min(\|y - y_d\|_2 + \alpha \|w_0\|_2) \quad (18)$$

$\alpha \gg 0$ پارامتر تنظیم سازی است. راه حل این مسئله به صورت زیر است:

$$w_0 = (V^T V + \alpha I)^{-1} V^T y_d \quad (19)$$

I یک ماتریس یک $n \times n$ است. هدف از بهینه سازی ELM، ماکزیمم کردن تابع *fitness* است:

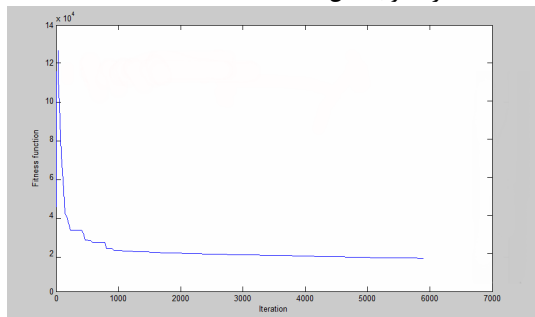
$$fitness = \frac{1}{1 + E_{rmse}(y, y_d)} \quad (20)$$

E_{rmse} میزان RMSE¹⁸ بین خروجی حقیقی و مقادیر خروجی تخمین زده شده توسط شبکه است. خطای تخمین $E_{rmse}(y, y_d)$ در مجموعه داده validation به دست می آید که هیچ تداخلی با مجموعه داده آموزشی ندارد و بصورت زیر تعریف می شود:

$$E_{rmse}(y, y_d) = \sqrt{\frac{1}{N} \sum_{k=1}^N [y(k) - y_d(k)]^2} \quad (21)$$

¹⁸ Root Mean Square Error

یک جمعیت کمترین fitness می‌باشد. این الگوریتم والدینی را انتخاب می‌کند که fitness بهتری داشته باشد. همانطور که نمودار نشان می‌دهد مقادیر fitness به سمت کمترین مقدار همگرا شده اند که این نشان دهنده عملکرد خوب مدل است.



شکل (3): نمودار همگرایی GA

شکل 4 روند آموزش مدل و مقایسه مقادیر واقعی و مقادیر پیش‌بینی شده برای مجموعه داده‌ها را نشان می‌دهد. همانطور که می‌توان دید مدل ارائه شده با تقریب خوبی توانسته است روندهای گذشته را شبیه سازی کرده و روندهای آینده را پیش‌بینی نماید. شکل 5 روند آموزش شبکه ELM که بهینه نشده است را نشان می‌دهد. لازم به ذکر است که از مجموع 44 نمونه، 36 نمونه برای آموزش شبکه و 8 نمونه در مرحله تست استفاده شده است. روش‌های مختلفی برای اندازه‌گیری دقت مدل‌های پیش‌بینی وجود دارد. رایج‌ترین آنها معیار ریشه میانگین مجذور خطای پیش‌بینی (RMSE) است. مدلی که کمترین مقدار ریشه میانگین مجذور خطای پیش‌بینی (فرمول 24) را داشته باشد به عنوان بهترین مدل پیش‌بینی انتخاب می‌شود. در این بررسی نیز از معیار مزبور استفاده شده است. این معیار به صورت زیر است:

$$RMSE = \left(\frac{T+H}{\sum_{i=T}^H (\hat{y}_t - y_t)^2 / h} \right)^{1/2} \quad (24)$$

که در آن \hat{y}_t و y_t به ترتیب مقدار پیش‌بینی شده و مقدار واقعی مورد پیش‌بینی هستند. در معادله شماره (1) پارامتر h تعداد مشاهدات مورد پیش‌بینی می‌باشد.

علاوه بر معیار فوق از معیار دیگری به نام معیار درصد میانگین خطای نسبی پیش‌بینی (Mean Absolute Percentage Error) استفاده شده است (فرمول 25). مزیت استفاده از شاخص درصد میانگین خطای نسبی این است که وابسته به مقیاس نیست و امکان مقایسه قدرت پیش‌بینی را برای سری‌هایی که دارای مقیاس متفاوت هستند فراهم می‌کند. این شاخص‌ها به صورت زیر تعریف می‌شوند:

$$MAPE = \left(\frac{T+h}{\sum_{i=T+1}^h |\hat{y}_t - y_t| / y_t} \right) / h \quad (25)$$

ناقص است و در بازه زمانی مدنظر ما وجود نداشت بنابراین برای این دو متغیر قیمت‌های جهانی آن مدنظر قرار گرفت.

از متغیرهای اقتصادی تورم و قیمت بازار دلار بصورت ماهانه در بازه زمانی فوق جمع‌آوری گردید. متغیرهای مربوط به شاخص‌های تولید صنعتی و تولید محصولاتی نظیر خودرو که مصرف‌کننده فولاد خام در کشور هستند بصورت ماهیانه در بازه زمانی فوق ناقص بود و عملاً امکان قرار دادن آنها در مدل‌سازی توابع عرضه و تقاضا وجود نداشت. با توجه به وجود متغیرهایی مانند تقاضای محصولات فولادی، ورق گرم و میلگرد که در صنایعی مانند تولید خودرو و ساخت‌وساز مورد استفاده قرار می‌گیرند، عملاً این شاخص‌ها می‌تواند نشان دهنده شاخص‌های تولید صنعتی بوده و بنابراین اطلاعات مورد نیاز ما را در خود خواهند داشت. با توجه به اینکه تعرفه واردات فولاد خام در سالهای فوق ثابت بوده است بنابراین در مدل‌سازی در نظر گرفته نمی‌شود. در مورد آهن اسفنجی و تولید آن نیز اطلاعات مورد نیاز بسیار ناقص بود و بنابراین در مدل‌سازی وارد نخواهند شد. در نهایت لیست متغیرهای مورد استفاده در این تحلیل و نماد مورد استفاده برای آنها در جدول 3 و 2 آورده شده است.

جدول (2): لیست متغیرهای مورد استفاده در این پژوهش

مصرف ظاهری	صادرات	واردات	مجموع تولید بخش خصوصی و دولتی	نام محصول
Consumption Steel	Export Steel	Import steel	Supply Steel	فولاد خام
Consumption SG	Export SG	Import SG	Supply SG	محصولات فولادی
Consumption Varaq	Export Varaq	Import Varaq	Supply Varaq	ورق گرم
Consumption Mil	Export Mil	Import Mil	Supply Mil	میلگرد

جدول (3): شاخص‌های مورد استفاده در این پژوهش

نام متغیر	نماد	نماد	نام متغیر
قیمت جهانی فولاد خام (\$/tonne)	Billet price	Inflation	نرخ تورم
قیمت جهانی قراضه (\$/tonne)	Steel Scrap	ERDollar	قیمت دلار

4-2 نتایج تجربی

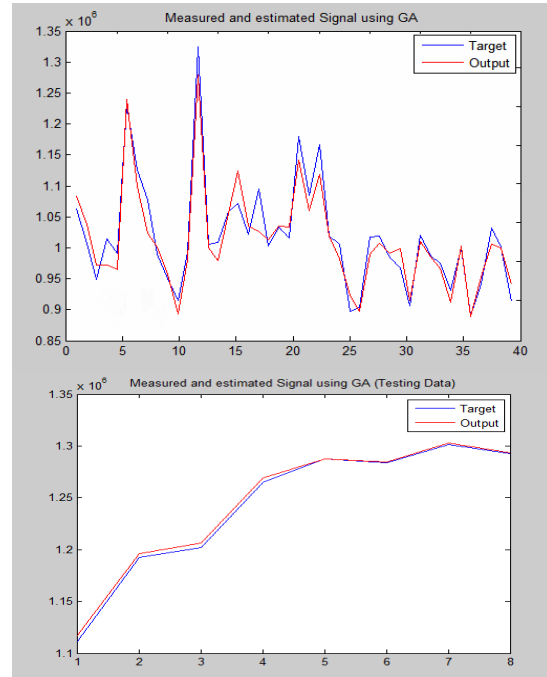
در این بخش نتایج حاصل از پیاده‌سازی مجموعه داده با 20 مشخصه و 44 به شبکه عصبی ارائه می‌شود. این شبکه عصبی با تابع انتقال سیگموئیدی و 19 نورون آموزش داده شده است. شکل 3 همگرایی الگوریتم ژنتیک روی دیتاست را نشان می‌دهد. مقدار fitness مقدار تابع برای افراد مختلف می‌باشد. به دلیل این که این جعبه ابزار وظیفه پیدا کردن کمترین fitness تابع را دارد، بهترین فیتنس برای

هرچه مقادیر این دو شاخص کوچکتر باشد، پیش بینی آرایه شده مطلوبتر خواهد بود. جدول 4 مقادیر این خطاها مربوط به شبکه ELM بهینه شده در مقایسه با سایر شبکه های عصبی کلاسیک را نشان می دهد. این شبکه یکی از نوترین و مشهورترین عضو خانواده شبکه های عصبی می باشند که تاکنون قابلیت های زیادی در حل مسائل مختلف از خود نشان داده است. آنچه از این جدول استنباط می شود آن است که خطای شبکه ELM بهینه در حد قابل قبولی است و این بدان معناست که شبکه به حد مناسبی از تعمیم دهی رسیده است و می توان از آن برای ورودی های متنوعی استفاده نمود. همان طور که از جدول 4 برمی آید خطای محاسباتی شبکه اصلی ELM بسیار بالاست و احتمالاً به این دلیل است که این شبکه در مواجهه با داده های با ابعاد کوچک دچار بیش برآزش می شود. این شبکه یکی از نوترین عضو خانواده شبکه های عصبی مصنوعی می باشد که تاکنون قابلیت های زیادی در حل مسائل مختلف از خود نشان داده است. مولفین در مطالعه ای دیگر به بررسی کارکرد این مدل در مواجهه با مجموعه داده های با ابعاد بسیار بالا پرداخته و نتایج بسیار خوبی بدست آورده اند. [31]

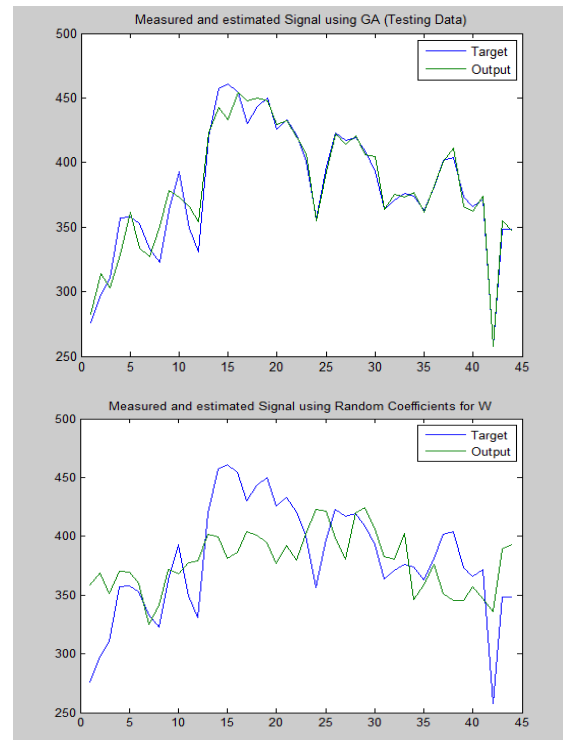
هم چنین محققان زیادی چگونگی بهینه سازی شبکه عصبی ELM به کمک الگوریتم ژنتیک را بررسی کرده اند که تعداد زیادی از این مطالعات به مسئله کلاس بندی می پردازد. برای مثال در این باره می توان به مطالعه ویمالا الکساندر و همکاران اشاره داشت [10].

آنها برای جلوگیری از گیرافتادن الگوریتم ژنتیک در بهینه های محلی از استراتژی نخه گرا²⁰ استفاده نمودند و نتایج حاصل از پیاده سازی شبکه عصبی در مسئله کلاس بندی را با سایر شبکه های مشابه مقایسه نمودند. این استراتژی با انتخاب افرادی با بهترین مقادیر fitness از عملگرهای crossover و جهش هم دوری می کند. اچانوبی و همکاران نیز به روش تقریباً مشابه از الگوریتم NSGA-ELM، یک الگوریتم ژنتیک چند هدفه به نحوی که همه اهداف با هم کمینه شوند، برای پردازش تصویر استفاده نموده اند [16].

پورنومول و همکاران [30] به کاربرد شبکه عصبی ELM-GA در مسئله پیش بینی سری زمانی پرداخته اند. الگوریتم آنان در هربار تکرار، هزینه محاسباتی (نرخ RMSE) را، با وزن های ورودی تصادفی، اندازه گیری نموده تا به مقادیر بهینه برسد در انتها دقت مدل ارائه شده نسبت به سایر شبکه های عصبی را با معیارهای RMSE و MSE سنجیده شده است. از طرفی همان طور که حل مشکلات Big Data از زمینه های تحقیقاتی مهم داده کاوی است، مجموعه داده با ابعاد پایین هم چالش مهمی محسوب می شود. چراکه همواره اطلاعات کامل و مفصلی از موضوع مورد بحث در دست نیست. آنچه در پژوهش حاضر مورد نظر نویسندگان بوده است بررسی کارایی شبکه عصبی ELM بهینه شده در مواجهه با مجموعه داده های با ابعاد کوچک است. شکل 6 زمان پردازش الگوریتم های مختلف، از لحاظ مدت زمان صرف شده برای یادگیری را نشان می دهد. می توان دید که سرعت یادگیری الگوریتم ELM سبب به

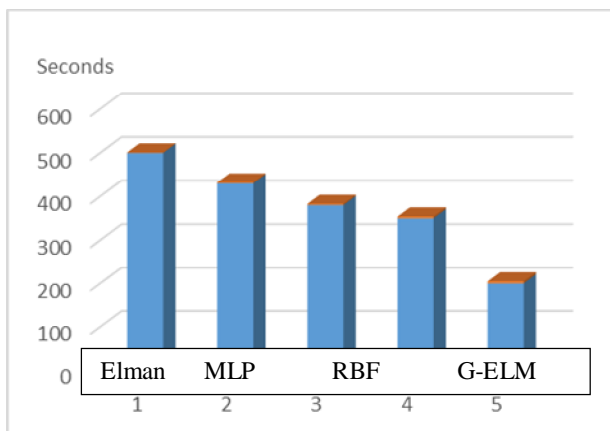


شکل (4): بالا: روند آموزش مدل در ELM بهینه شده، پایین: مقادیر واقعی و پیش بینی شده عرضه فولاد خام توسط مدل



شکل (5): روند آموزش شبکه ELM مربوط به داده های عرضه ی فولاد (بالا: با استفاده از الگوریتم ژنتیک و پایین: بدون استفاده از الگوریتم ژنتیک)

²⁰ Elitist Strategy



شکل (6): زمان پردازش الگوریتم‌های مختلف

برای ارزیابی مدل پیشنهادی، مجموعه داده عرضه و تقاضای ماهانه فولاد در کشور در بازه زمانی دی ماه 88 تا مرداد ماه 92 جمع‌آوری و بر روی مدل پیاده‌سازی شد. سپس مقایسه عملکرد بین این شبکه با شبکه‌های عصبی کلاسیک، از لحاظ سنجش کارایی و سرعت یادگیری انجام شد. در عمل، سرعت یادگیری ELM بیش از دو برابر سریع‌تر از شبکه‌های عصبی است. پس ELM زمان محاسباتی را به شدت کاهش می‌دهد از نظر تطابق بین مقادیر مشاهده شده و پیش‌بینی شده، بهبود چشمگیری در مدل ELM برای بدست آوردن مقادیر ماهیانه فولاد نسبت به شبکه‌های عصبی بوده است. مقادیر پیش‌بینی شده توسط ELM به مقادیر واقعی بسیار نزدیک بوده است. با توجه به مقادیر خطای MAPE RMSE، می‌توان گفت عملکرد آموزش و دقت پیش‌بینی در GELM به مراتب بهتر از ELM و سایر شبکه‌های عصبی است. بنابراین، این شبکه عصبی نمایش دقیق‌تری از تقاضای ماهیانه داشته است.

منابع و مأخذ

- [1] آذربایجانی، کریم، رضایی، محمدرضا، (1380)، بررسی پارامترها و کشش پذیری تقاضای فولاد کشور در دوره 78-46، فصلنامه پژوهش‌های اقتصادی ایران، 3(8)، 101-114.
- [2] اصغری اسکونی، محمدرضا، (1381)، کاربرد شبکه‌های عصبی در پیش‌بینی سری‌های زمانی، فصلنامه پژوهش‌های اقتصادی ایران، 4(12)، 69-96.
- [3] آقاجانی، حسین علی، صفایی، بهزاد، باصولی، امیر، (1392)، شناسایی و اولویت بندی راهکارهای بهبود مصرف انرژی در صنعت با استفاده از تکنیکهای تصمیم‌گیری چند معیاره (مطالعه موردی فولاد آلیاژی ایران)، تحقیق در عملیات در کاربردهای آن، 10(2)، 1-21.
- [4] آقای، کیومرث، پورمیری، بهروز، (1385)، پیش‌بینی روند قیمت فولاد با استفاده از شبکه‌های عصبی مصنوعی (ANN) و مقایسه نتایج آن با روش ARIMA، فصلنامه اقتصاد مقداری، 8، 144.
- [5] حافظی، رضا، شهرابی، جمال، هداوندی، اسماعیل، (1392)، توسعه مدل ترکیبی هوشمند برای پیش‌بینی بازار سهام تهران، تحقیق در عملیات در کاربردهای آن،

شبکه‌های عصبی دیگر بیشتر بوده است. اما عملاً وقتی از الگوریتم ژنتیک استفاده کنیم اگرچه دقت بهبود می‌یابد اما زمان یادگیری افزایش پیدا می‌کند.

5- نتیجه‌گیری

این پژوهش، رویکرد نوینی براساس الگوریتم ژنتیک برای بهبود شبکه عصبی ELM در مسئله پیش‌بینی تقاضای فولاد خام کشور ارائه داده است. این شبکه ویژگی‌های زیر را دارد:

- برخلاف شبکه‌های عصبی سنتی، نیاز به تنظیم مکرر پارامترها نیست و به این دلیل سرعت آموزش بالاتر می‌رود.
- با استفاده از الگوریتم ژنتیک و گزینش بهترین ورودی‌ها، از میل به شبکه‌های ELM به ساخت تعداد نورونهای زیاد ممانعت به عمل آمد و اندازه شبکه کاهش پیدا کرد.
- نیاز به انتخاب تعداد نورون‌های لایه مخفی با روش آزمون و خطا نیست و ورودی‌های مرتبط به صورت خودکار گزینش می‌شوند.
- عملکرد آموزش و دقت پیش‌بینی در ELM بهینه شده به مراتب بهتر از ELM و سایر شبکه‌های عصبی کلاسیک است.

جدول (4): مقایسه نرخ خطای آموزشی در شبکه‌های عصبی مختلف

شبکه عصبی مورد استفاده	تعداد نمونه‌ها	RMSE	MAPE
ELM بهینه	35 داده آموزشی و 8 داده آزمایشی	0/13256	0/0565
ELM	35 داده آموزشی و 8 داده آزمایشی	1/00403	0/9114
RBF	35 داده آموزشی و 8 داده آزمایشی	0/4899	0/3080
ELMAN	35 داده آموزشی و 8 داده آزمایشی	0/8543	0/1025
MLP	35 داده آموزشی و 8 داده آزمایشی	0/9322	0/0691

- [27] Minhas, R., Baradarani, A., Seifzadeh, S., Wu, Q.M.J., (2010) **Human Action Recognition Using Extreme Learning Machine Based on Visual Vocabularies**, Neurocomputing, 73, 1906–1917.
- [28] Mohammed, A.A., Wu, Q.M.J., Sid-Ahmed, M.A., (2010) **Application of Wave Atoms Decomposition and Extreme Learning Machine for Fingerprint Classification**, Lecture Notes Computer Science, 6112, 246–256
- [29] Paul, C., (2000), **Future Trends in Japanese Steel Consumption**, Resources Policy, 26, 103–114.
- [30] Purnomol, D. M. J., Purbarani, S. C., Wibisonol, A., Hendrayanti, D. et al. (2015), **Genetic Algorithm Optimization for Extreme Learning Machine Based Micro algal Growth Forecasting of Chlamydomonas**, ICACSI 2015, IEEE, 243-248.
- [31] Rezaeenour, J., Yari Eili, M., Roozbahani, Z., Ebrahimi, M., (2016), **Prediction of Protein Thermostability by an Efficient Neural Network Approach**, Journal of Health Management and Informatics, 3(4), 102-110.
- [32] Robert, M.C., (1990), **Prediction Metal Consumption: the Case Study of US Steel**, Resource policy, 16(1), 56-73.
- [33] Shahrazi, J., Asadi, E., (2013), **Developing a Hybrid Intelligent Model for Forecasting Problems: Case Study of Tourism Demand Time Series**, Knowledge-based Systems, 43, 112-122
- [34] Xin, L., Findlay, C., (1985), **China Iron and Steel Industry Policy: Implications for Australia**, Canberra: Research School of Pacific Studies, Australian National University.
- [35] Yin, X., Chen, W., (2013), **Trends and Development of Steel Demand in China: A Bottom-up Analysis**, Resources Policy, 38, 407–415.
- [36] Yu, Y., Choi, T.M., Hui, C.L., (2012), **An Intelligent Quick Prediction Algorithm with Applications in Industrial Control and Loading Problems**, IEEE Transactions on Automation Science and Engineering, 9(2), 276–287.
- [6] خدابنده، مهدی، حجازی، سیدرضا، راستی برزکی، مرتضی، (1392)، **یک الگوریتم ژنتیک برای مسئله زمانبندی یکپارچه تولید و توزیع با در نظر گرفتن مسیربایی در زنجیره تامین**، نشریه پژوهش‌های مهندسی صنایع در سیستم‌های تولید 1(2)، 181-167.
- [7] خسروشاهی، حسین، معطرحسینی، سید محمد، مرجانی، محمدرضا، (1393)، **اندازه گیری اثر شلاق چرمی در یک زنجیره تامین خطی سه سطحی با استفاده از روش میانگین متحرک برای برآورد تقاضا**، نشریه پژوهش‌های مهندسی صنایع در سیستم‌های تولید، 2(4)، 21-37.
- [8] گودرزی، حسین، (1386)، **پیش‌بینی تقاضای فولاد خام ایران در سال 1400**، دو فصلنامه برنامه و بودجه، 12(4)، 209-232.
- [9] میرفخرالدینی، سید حیدر، بابایی میبدی، حمید، مروتی شریف آبادی، علی، (1392)، **پیش‌بینی مصرف انرژی ایران با استفاده از مدل ترکیبی الگوریتم ژنتیک- شبکه عصبی مصنوعی و مقایسه آن با الگوهای سنتی**، پژوهش‌های مدیریت در ایران، 17(2)، 196-222.
- [10] Alexander, V., Annamalai, P., (2015), **An Elitist Genetic Algorithm Based Extreme Learning Machine**, *Computational Intelligence, Cyber Security and Computational Models*, 301-309.
- [11] Baradarani, A., Wu, Q.M.J., Ahmadi, M., (2013), **An Efficient Illumination Invariant Face Recognition Framework via Illumination Enhancement and DD-DTC WT Filtering**, Pattern Recognition, 46(1), 57–72
- [12] Benoit, F., van Heeswijk, M., Miche, Y., Verleysen, M., Lendasse, A., (2013), **Feature Selection for Nonlinear Models with Extreme Learning Machines**, Neurocomputing, 102, 111–124.
- [13] Deo, R.C., Sahin, M., (2015), **Application of Extreme Learning Machine Algorithm for the Prediction of Monthly Effective Drought Index in Eastern Australia**, Atmospheric Research, 153, 512-525.
- [14] Decherchi, S., Gastaldo, P., SDahiya, R., Valle, M., Zunino, R., (2011), **Tactile Data Classification of Contact Materials Using Computational Intelligence**, IEEE Transactions on Robotics, 27(3), 635–639.
- [15] Doquire, G., Verleysen, M., (2012), **Feature Selection with Missing Data Using Mutual Information Estimators**, Neurocomputing, 90, 3–11.
- [16] Echanobe, J., del Campo, I., Martínez, v., Basterretxea, K., (2017), **Genetic Algorithm-based Optimization of ELM for On-line Hyperspectral Image Classification**, IJCNN, IEEE, 4202-4207.
- [17] Grigorievskiy, A., Miche, Y., Ventelä, A.M., Séverin, E., Lendasse, A., (2014), **Long-Term Time Series Prediction Using OP-ELM**, Neural Networks, 51, 50–56.
- [18] Haupt, R. L., Haupt, S. E., (2004), **Practical Genetic Algorithms**, 2nd edn with CD-ROM. John Wiley & Sons, New York, NY.
- [19] Honerkamp, J., Weese, J., (1990), **Tikhonov's regularization method for ill-posed problems**, Continuum Mechanics and Thermodynamics, 2, 17–30.
- [20] <http://www.steelonthenet.com/semi-finished-prices.php>
- [21] Huang, G.B., Yu Zhu, Q., Siew, Ch.Kh., (2006), **Extreme Learning Machine: Theory and Applications**, Neurocomputing, 70, 489–501.
- [22] Huang, G.B., (2012), **Extreme Learning Machine for Regression and Multiclass Classification**, IEEE Trans. Systems, Man, and Cybernetics, Part B: Cybernetics, 42(2), 513–529.
- [23] Huang, G., Huang, G.B., Song, S., You, K., (2015), **Trends in Extreme Learning Machines: A Review**, Neural Networks, 61, 32–48.
- [24] Huang, G.B., Wang, D.H., Lan, Y., (2011), **Extreme Learning Machines: a Survey**, International Journal of Machine Learning & Cyber. 2, 107–122.
- [25] Liu, Xu, Gao, Ch., Li, P., (2012), **A Comparative Analysis of Support Vector Machines and Extreme Learning Machines**, Neural Networks, 33, 58–66.
- [26] Matias, T., Araújo, R., Henggeler Antunes, C., Gabriel, D., (2013)