

# ارائه روش ترکیبی پیش پردازش داده‌ها در ماشین بردار رگرسیون جهت پیش‌بینی کیفیت گازوییل پالایش شده

پژوهش نفت

سال بیست و سوم  
شماره ۷۵

صفحه ۱۱۶-۱۰۲ ۱۳۹۲

تاریخ دریافت مقاله: ۹۲/۲/۲۳

تاریخ پذیرش مقاله: ۹۲/۶/۹

سعید شگری<sup>۱</sup>، محمد تقی صادقی<sup>۱\*</sup> و مهدی احمدی مروست<sup>۲</sup>

۱- دانشگاه علم و صنعت ایران، دانشکده مهندسی شیمی

۲- پژوهشگاه صنعت نفت، پژوهشکده توسعه فرآیند و فناوری تجهیزات

sadeghi@iust.ac.ir

سنجش دقت پیش‌بینی مدل ماشین بردار رگرسیون استفاده گردید. مقدار AARE برابر ۰/۰۵۳ به دست آمد که نشان‌دهنده دقت بالای مدل در پیش‌بینی غلظت گوگرد خروجی از رآکتور می‌باشد.

واژه‌های کلیدی: آنالیز موجک، حسگر مجازی، ماشین بردار پشتیبان، تلفیق داده‌ها، پیش‌بینی کیفیت

## مقدمه

استفاده از حسگرهای مجازی (نرم‌افزاری) برای تعیین کیفیت محصول و محاسبه متغیرهایی که امکان اندازه‌گیری آنها مشکل است، بسیار مورد توجه می‌باشد. طراحی حسگرهای مجازی که بر مبنای داده‌های ورودی کار می‌کنند، نیازمند داده‌های صحیح و قابل اعتماد می‌باشند. اندازه‌گیری‌های صحیح و معتبر نقش تعیین‌کننده‌ای در پیش‌بینی حاصل از آنالیزهای نرم‌افزاری بر عهده دارند. در بهره‌برداری از فرآیند گوگردزایی با هیدروژن، تعیین مقدار غلظت گوگرد خروجی یکی از شاخص‌های مهم در کیفیت محصول می‌باشد. به این منظور از حسگرها و آنالیزهای سخت‌افزاری استفاده می‌گردد، ولی این آنالیزها معمولاً گران بوده، نیاز به کالیبره شدن دارند و از لحاظ نگهداری هزینه بر می‌باشند.

## چکیده

از آنجا که دقت داده‌ای اندازه‌گیری شده فرآیندی در پیش‌بینی کیفیت محصولات بسیار مهم است، در این تحقیق بر روی پیش‌پردازش داده‌ها تمرکز گردید. برای این منظور حسگر مجازی برای تعیین کیفیت گازوییل خروجی از پیلوت تصفیه هیدروژنی طراحی شد. طراحی حسگر مجازی بر اساس یکی از روش‌های جدید یادگیری ماشین به نام ماشین بردار رگرسیون انجام گردید. برای پیش‌پردازش داده‌ها از تکنیک ترکیبی به صورت پشت سر هم متشکل از آنالیز موجک و کوانتیزاسیون برداری به منظور حذف خطاهای تصادفی، متراکم‌سازی داده‌ها و چشم‌پوشی از داده‌هایی که شباهت کمتری به سایر داده‌ها دارند، استفاده گردید. روش‌های متفاوتی از آنالیز موجک برای حذف خطاهای تصادفی به کار برده شد و بهترین روش انتخاب گردید. آزمایشات حذف خطاهای تصادفی با استفاده از آنالیز موجک با تابع پایه هار و دایجیز و با الگوریتم‌های انتخاب آستانه  $\text{Heursure}$ ،  $\text{Rigrsure}$ ،  $\text{Minimaxi}$  و  $\text{Sqtwolog}$  انجام شد. مقایسه نتایج نشان داد که روش  $\text{Db4}$  به همراه روش آستانه‌گیری  $\text{Rigrsure}$  بهترین نتایج حذف خطا را به دنبال دارد. با استفاده از این روش مقدار عددی  $\text{AARE}$  و  $\text{RMSE}$  نسبت به انواع دیگر تابع موجک بهتر است. همچنین، معیار عملکردی  $\text{AARE}$  برای

حالی که مدل ماشین بردار پشتیبان به خوبی تعمیم پذیر است [۱۰]. ویژگی مهم ماشین بردار پشتیبان این است که برخلاف الگوریتم‌های کلاسیک و رگرسیون‌های خطی که به وسیله می‌نیمم کردن قدر مطلق خطا یا توان دوم خطا عمل می‌کنند، آنها ریسک عملیاتی را می‌نیمم می‌کنند. ماشین بردار پشتیبان بر خلاف شبکه‌های عصبی با مشکل گیر افتادن در می‌نیمم‌های محلی تابع خطا مواجه نمی‌باشد [۱۱-۱۳]. همچنین با استفاده از هسته‌های غیر خطی قادر به تصمیم‌گیری غیر خطی نیز می‌باشد. انتخاب هسته‌های مناسب برای ماشین بردار پشتیبان، منجر به برتری آن نسبت به سایر رویکردهای مبتنی بر تصمیم‌گیری خطی شده است. شروع استفاده از ماشین‌های بردار پشتیبان به زمانی برمی‌گردد که نظریه یادگیری آماری در سال ۱۹۷۹ در روسیه به وسیله وپنیک، بوسیر و همکارانش گسترش یافت. ماشین‌های بردار پشتیبان به صورتی که امروزه مورد استفاده قرار می‌گیرند، برای اولین بار در سال ۱۹۹۲ در کنفرانس COLT به وسیله وپنیک و همکارانش معرفی گردید. آنها در سال ۱۹۹۵ ماشین‌های بردار پشتیبان را برای کاربرد در مسائل رگرسیون توسعه دادند [۱۴ و ۱۵].

بنابراین، در این مطالعه برای طراحی حسگر مجازی از ماشین بردار پشتیبان بهره گرفته شد. از آنجایی که صحت داده‌های ورودی در پیش‌بینی حسگر مجازی بسیار مهم می‌باشد، لازم است داده‌ها قبل از ورود به این حسگرها عاری از انواع نویزها و خطاهای تصادفی باشند. به عبارت دیگر داشتن داده‌های اندازه‌گیری شده صحیح و قابل اعتماد یکی از ابزارهای ضروری برای طراحی حسگرهای نرم‌افزاری است و با استفاده از اطلاعات معتبر می‌توان پیش‌بینی‌ها را با ضریب اطمینان بالاتر انجام داد. داده‌های فرآیند که از سیستم‌های اطلاعاتی به دست می‌آیند، بیشتر اوقات غیرقابل اطمینان و غیردقیق هستند. خطاهای تصادفی و نویزها در اندازه‌گیری‌ها، باعث می‌شود محاسبات بر مبنای داده‌های غیرمطمئن انجام شده و خطاها گسترش پیدا کند. خطاهای تصادفی در فرآیند ناشی از نوسانات و یا آشفتگی‌های مربوط به شرایط کنترل نشده می‌باشد که باعث غیر قابل اطمینان بودن اندازه‌گیری‌های انجام شده، می‌گردد. اندازه‌گیری‌هایی که شامل این نوع از خطاها

در کاربردهای عملی متداول است که حسگر مجازی (نرم‌افزاری) در کنار حسگرهای معمولی استفاده شود بدین ترتیب نقش حسگر مجازی کاهش هزینه‌های نگهداری خواهد بود. اخیراً، کار با حسگرهای مجازی (نرم‌افزاری) که بر اساس داده‌های ورودی طراحی می‌شوند، بسیار اهمیت یافته است [۱]. این حسگرها دارای انواع مختلفی هستند که طراحی آنها بر اساس مدل یا داده‌های فرآیندی می‌باشد. در این تحقیق برای پیش‌بینی غلظت گوگرد خروجی از حسگر مجازی که بر اساس داده‌های ورودی کار می‌کند، استفاده شده است. از آنجا که با پیشرفت کامپیوترهای پرسرعت و تجهیزات اندازه‌گیری و داده‌برداری در فرآیندهای شیمیایی، حجم وسیعی از داده‌ها در دسترس می‌باشد، استفاده از این روش مورد توجه بسیاری از محققان قرار گرفته است [۲ و ۳].

در آمار کلاسیک روش‌های طبقه‌بندی و رگرسیون<sup>۱</sup> بر پایه فرضیات محدود کننده‌ای بنا شده که در آن مدل‌های توزیع احتمال و یا توابع چگالی احتمال، معلوم هستند. متأسفانه بسیاری اوقات، در عمل اطلاعات کافی درباره توزیع احتمال متغیرهای مورد مطالعه در دسترس نیست. در چنین مواقعی به روش‌هایی نیاز داریم که بدون دانستن توزیع احتمال، به خوبی عمل کنند. همچنین در بیشتر مطالعات با اطلاعاتی در فضاهایی با ابعاد بالا مواجه هستیم. برای استفاده از روش‌های کلاسیک آماری در چنین شرایطی، نیازمند نمونه‌هایی با حجم بالا می‌باشیم که ممکن است در عمل فراهم کردن آن میسر نباشد. یکی از روش‌هایی که برای حل چنین مشکلاتی مورد استفاده قرار می‌گیرد، استفاده از الگوریتم‌های یادگیری ماشین است.

ماشین بردار پشتیبان یک تکنیک جدید از روش‌های یادگیری ماشین است و می‌توان محبوبیت کنونی آن را با محبوبیت شبکه‌های عصبی در دهه‌های گذشته مقایسه کرد. این تکنیک به تازگی توجه محققان فرآیند را در خصوص طراحی حسگرهای مجازی به خود جلب کرده است [۴-۸].

ماشین بردار پشتیبان در پیش‌بینی‌ها نسبت به شبکه‌های عصبی از دقت بالاتری برخوردار است [۹]. از طرفی تعیین وضعیت شبکه و قابلیت تعمیم شبکه عصبی برای وظایف مدل‌سازی/یادگیری هنوز به خوبی حل نشده است، در

نقاط آموزشی را به صورت  $[x_i, y_i]$  و بردار ورودی  $x_i \in \mathbb{R}^n$  تعریف کنیم، در حالتی که داده‌ها به صورت خطی قابل تفکیک باشند، معادله به صورت زیر است:

$$y = f(x) = \text{sign}[\sum_{i=1}^N y_i a_i(x_i, x) + b] \quad (5)$$

که در آن  $y$  خروجی معادله و  $y_i$  ارزش طبقه نمونه آزمایشی  $x_i$  می‌باشد. بردار  $x = (x_1, x_2, \dots, x_n)$  نشان دهنده یک داده ورودی و بردارهای  $x_i, i=1, 2, \dots, N$  بردارهای پشتیبان هستند. اگر داده‌ها به صورت خطی قابل تفکیک نباشند، می‌توان با اعمال پیش پردازش، نمونه‌ها را به فضایی بالاتر برد. در این حالت رابطه ۵ به رابطه ۶ تغییر می‌یابد:

$$y = f(x) = \text{sign}[\sum_{i=1}^N y_i a_i K(x, x_i) + b] \quad (6)$$

تابع  $K(x, x_i)$  تابع کرنلی<sup>۴</sup> است که برای ایجاد ماشین‌هایی با انواع مختلفی از سطوح غیر خطی در فضای داده‌ها، ضرب‌های داخلی تولید می‌کند. برای مدل رگرسیون ماشین بردار پشتیبان از کرنل‌های مختلفی استفاده می‌شود که عبارتند از خطی، درجه دوم، گوسی و چند جمله‌ای. معمولاً تابع کرنل گوسی شعاعی<sup>۵</sup> برای پیش‌بینی عملکرد بهتری دارد [۱۸]. معادله این تابع کرنل به صورت زیر است.

$$K(x, y) = \exp\left(-\frac{\|x_i - y_j\|^2}{2\sigma^2}\right) \quad (7)$$

در ساخت یک مدل کارآمد ماشین بردار پشتیبان، پارامترهای مدل باید با استفاده از یک روش بهینه‌سازی به طور دقیق محاسبه گردند. این پارامترها عبارتند از:

- نوع تابع کرنل
- پارامتر تابع کرنل  $\sigma^2$
- پارامتر تنظیم کننده  $C$
- پارامتر دقت  $\epsilon$  مربوط به حداکثر خطا در ناحیه  $\epsilon$ -Insensitive یا حداکثر شعاع لوله واقع در اطراف تابع رگرسیون حذف خطاهای تصادفی با استفاده از آنالیز موجک

تکنیک فیلترینگ داده‌ها به مفهوم تخمین مقادیر واقعی از مقادیر اندازه‌گیری شده و بر مبنای برخی اطلاعات اضافی (برای مثال مدل فرآیند) می‌باشد.

می‌باشند، در قیود حاکم بر سیستم صدق نکرده و از آن انحراف نشان می‌دهند.

## روش کار

### مدل رگرسیون ماشین بردار پشتیبان<sup>۱</sup>

مدل‌های ماشین‌های بردار پشتیبان به دو گروه عمده الف) مدل طبقه‌بندی ماشین بردار پشتیبان<sup>۲</sup> و ب) مدل رگرسیون ماشین بردار پشتیبان تقسیم بندی می‌شوند. از مدل طبقه‌بندی ماشین بردار پشتیبان جهت حل مسائل طبقه‌بندی داده‌هایی که در کلاس‌های مختلف قرار می‌گیرند استفاده می‌گردد و مدل رگرسیون ماشین بردار پشتیبان در حل مسائل پیش‌بینی کاربرد دارد.

همان‌گونه که بیان شد، ماشین بردار پشتیبان مبتنی بر می‌نیم کردن ساختاری ریسک<sup>۳</sup> می‌باشد که از تئوری آموزش آماری گرفته شده است [۱۶]. وپنیک برای کاربرد ماشین‌های بردار پشتیبان در مسائل رگرسیون از تابع خطایی استفاده کرد که خطاهایی که در یک  $\epsilon$ -Insensitive به نام فاصله معین از مقادیر واقعی می‌باشند، را نادیده می‌گیرد [۱۷]. این تابع به صورت رابطه زیر تعریف می‌شود:

$$L(y, f(x, a)) = |y - f(x, a)|_{\epsilon} = \begin{cases} 0 & \text{for } |y - f(x, a)| \leq \epsilon \\ |y - f(x, a)| - \epsilon & \text{if } |y - f(x, a)| > \epsilon \end{cases} \quad (1)$$

این تابع خطا مقادیر خطای کمتر از  $\epsilon$  را در نظر نمی‌گیرد.

مسئله تقریب یک مجموعه از داده‌های زیر را در نظر می‌گیریم:

$$D = \{(x^1, y^1), \dots, (x^l, y^l)\}, \quad (x \in \mathbb{R}^n, y \in \mathbb{R}) \quad (2)$$

تابع رگرسیون به وسیله تابع زیر تخمین زده می‌شود.

$$f(x) = \langle w, x \rangle + b \quad (3)$$

که  $\langle \cdot \rangle$  ضرب داخلی می‌باشد. تابع بهینه رگرسیون به وسیله می‌نیم تابع زیر بیان می‌گردد.

$$\Phi(w, \xi) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l (\xi_i^- + \xi_i^+) \quad (4)$$

$$\text{subject to } \begin{cases} y_i - (\langle w, x_i \rangle + b) \leq \epsilon + \xi_i \\ (\langle w, x_i \rangle + b) - y_i \leq \epsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \end{cases}$$

که  $C$  یک مقدار از پیش تعیین شده بوده و  $\xi_i^+, \xi_i^-$  متغیرهای شل هستند که قیود بالا و پایین خروجی سیستم را تعیین می‌کنند. اگر داده‌ها به صورت خطی مجزا از هم باشند، یک سطح بهینه که داده‌ها را بدون خطا و با حداکثر فاصله میان صفحه و نزدیک‌ترین نقاط آموزشی (بردارهای پشتیبان) تفکیک می‌نماید، آموزش می‌دهد. اگر

1. Support Vector Regression (SVR)  
2. Support Vector Classification  
3. Structural Risk Minimization (SRM)  
4. Kernel Function  
5. Gaussian Radial Basis Kernel Function (RBF)

بین روش‌های ذکر شده استفاده گردیده است.

تبدیل موجک پیوسته<sup>۴</sup> به صورت زیر است:

$$W_{a,b}(x) = \int_{-\infty}^{+\infty} x(t) \psi_{a,b}^*(t) dt \quad (۸)$$

که تابع موجک مادر<sup>۵</sup> به وسیله رابطه زیر تعریف می‌شود:

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}} \psi\left(\frac{t-b}{a}\right) \quad (۹)$$

$a$  یک فاکتور مقیاس و  $b$  یک فاکتور انتقال است [۲۲]. پس از گسسته‌سازی، موجک به صورت زیر در می‌آید [۲۳]:

$$\psi_{j,k}(t) = a_0^{-\frac{j}{2}} \psi(a_0^{-j}t - kb_0) \quad (۱۰)$$

فیلترهای موجک دارای انواع مختلفی می‌باشد که از مهم‌ترین آنها می‌توان به فیلترهای موجک دایبچز<sup>۶</sup> با مرتبه‌های مختلف و فیلترهای موجک هار<sup>۷</sup> اشاره کرد.

متداول‌ترین خانواده موجک برای تجزیه و تحلیل سیگنال‌های موجک دایبچز می‌باشد. انتخاب تابع موجک مناسب نقش مهمی در دست‌یابی به عملکرد بهینه در پردازش سیگنال دارد [۲۴ و ۲۵].

#### کوانتیزاسیون برداری

در فرآیندها معمولاً با حجم زیادی از داده‌ها مواجه می‌شویم که این مسأله می‌تواند زمان آموزش را افزایش دهد. برای حل این مشکل می‌توان از تکنیک متراکم‌سازی داده‌ها با روش کوانتیزاسیون برداری<sup>۸</sup> استفاده کرد.

جایگزین نمودن حجم انبوه داده‌های موجود در فضای ورودی، با تعداد محدودی از داده‌ها، موسوم به بردارهای کتاب کد صورت می‌گیرد. روش کوانتیزاسیون برداری جهت کلاس‌بندی بردارهای آزمایشی، سعی در یافتن نزدیک‌ترین بردار (بردارها) کتاب کد به بردار ورودی دارد. بنابراین می‌توان گفت اطلاعات حاصل از روش کلاس‌بندی فوق، نشان‌دهنده فاصله بین بردار آموزشی با نزدیک‌ترین بردار (یا  $k$  نزدیک‌ترین بردار) کتاب کد می‌باشند [۲۶]. جهت کلاس‌بندی صحیح نمونه‌ها، لازم است فاصله اقلیدسی<sup>۹</sup> بین بردارهای آموزشی و نزدیک‌ترین بردارهای کتاب کد حداقل شود.

1. Statistic-based filtering
2. Outlier
3. Thresholding
4. Continuous Wavelet Transform (CWT)
5. Mother Wavelet
6. Daubechies Wavelets Filter (Dbn)
7. Har Wavelets
8. Vector Quantization (VQ)
9. Euclidean Distance

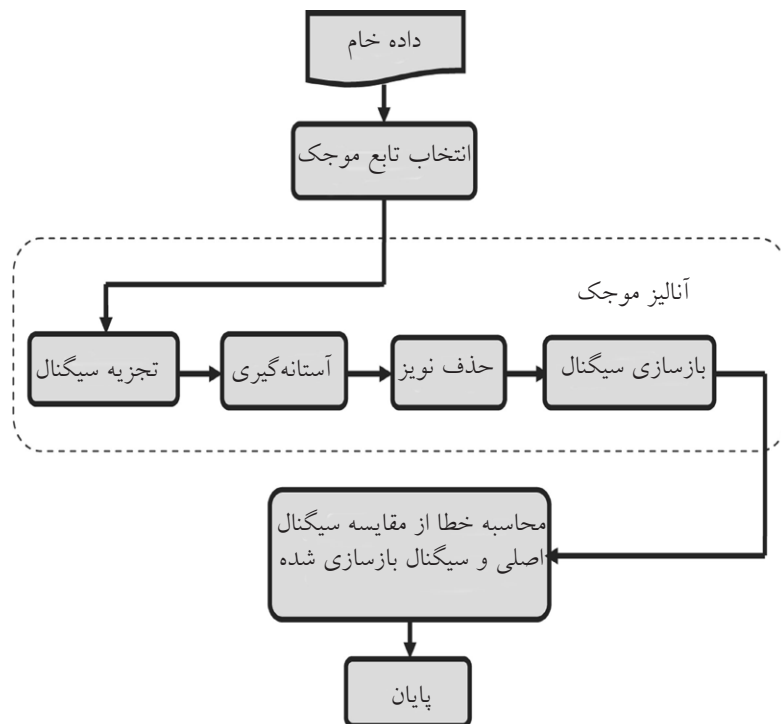
معمولاً فیلترینگ اندازه‌گیری‌های به دست آمده از یک فرآیند بر مبنای دو روش فیلترینگ بر مبنای مدل و فیلترینگ آماری<sup>۱</sup> می‌باشد [۱۹].

در این تحقیق از فیلتر آنالیز موجک در عملیات نویززدایی از داده‌ها و حذف خطاهای تصادفی استفاده شده است. آنالیز موجک یکی از دستاوردهای نسبتاً جدید ریاضیات محض می‌باشد که امروزه به یکی از مهم‌ترین ابزارها در زمینه پردازش سیگنال تبدیل شده است. تبدیل موجک سیگنال را از فضای آرگومان به فضای فرکانس تبدیل می‌کند [۲۰]. آنالیز موجک با در نظر گرفتن مقدار واریانس نویز متغیرهای اندازه‌گیری شده را در طول زمان نویززدایی می‌کند. بنابراین، می‌توان از آن در پیش پردازش سیستم‌های پویا نیز بهره برد. این آنالیز قادر است با حذف اغتشاشات بزرگ<sup>۲</sup>، داده‌ها را یکنواخت و هموار نماید [۲۱].

باید توجه داشت که ساده‌ترین و ابتدایی‌ترین روش‌ها در حذف نویز، استفاده از فیلترهای خطی می‌باشد. ولی این‌گونه فیلترها باعث بدنمایی سیگنال می‌شوند. بنابراین به طور معمول مورد استفاده قرار نمی‌گیرند. با پیشرفت‌هایی که در الگوریتم‌های حذف نویز با استفاده از تبدیل موجک صورت گرفته، این مشکل مرتفع گردیده است و تبدیل موجک علاوه بر حذف نویز به صورت مناسب، خواص اصلی سیگنال را نیز حفظ می‌کند.

مراحل اصلی کاهش یا حذف نویز در سیستم‌های تبدیل موجک شامل تجزیه سیگنال (توسط تبدیل موجک)، حذف نویز (حذف ضرایب کوچک در تبدیل حاصل موسوم به آستانه‌گیری<sup>۳</sup>) و نهایتاً بازسازی سیگنال می‌باشد.

در این تحقیق پس از ذخیره داده‌های پایلوت گوگردزدایی در دیسک سخت، این داده‌ها در نرم‌افزار *MATLAB* فراخوانی شده و هر متغیر به صورت برداری جدا در سیستم ذخیره می‌شود. پس از آن هر بردار توسط خانواده‌ای از موجک‌ها تجزیه گردیده و توسط الگوریتم‌های مختلف حذف آستانه‌ای، ضرایب موجک‌ها مورد آنالیز قرار می‌گیرد. سیگنال مورد نظر دوباره توسط همان خانواده از موجک‌ها بازسازی می‌شود. به منظور بررسی میزان عملکرد روش به کار رفته از معیار خطاهای *RMSE* و *AARE* برای مقایسه



شکل ۱- دیاگرام الگوریتم آنالیز موجک

گفت در صورتی که داده‌های اولیه به خوبی انتخاب شوند، بازده سیستم ترکیبی در پیش‌بینی تغییر چندانی نداشته و در بیشتر مواقع افزایش می‌یابد.

#### تست‌های مختلف پایلوتی و اخذ داده‌های فرآیندی

در پایلوت گوگردزدایی از سوخت دیزل، خوراک محتوی گوگرد با غلظت  $7200 \text{ ppm}$ ، وارد راکتور شده و پس از گوگردزدایی در حضور هیدروژن در یک راکتور با بستر قطره‌ای، میزان گوگرد کاهش می‌یابد. آزمایشات گوگردزدایی از دیزل توسط کاتالیست کبالت/مولیبدن بر پایه آلومینا انجام می‌شود. برای فرآیند *HDS* دو نوع خوراک مورد نیاز است. خوراک نوع اول، به منظور فعال‌سازی کاتالیست مورد استفاده قرار می‌گیرد و خوراک نوع دوم، خوراکی است که فرآیند جداسازی ترکیبات گوگردی بر روی آن انجام می‌شود.

خوراک نوع اول شامل دی سولفید کربن، سولفید هیدروژن و دی متیل دی سولفید می‌باشد و خوراک نوع دوم شامل نفتا، دیزل و هر ترکیب هیدروکربنی دیگر از مشتقات نفت خام می‌باشد. شماتیک و مشخصات این فرآیند در شکل ۲ نشان داده شده است.

جهت کلاس‌بندی دنباله ورودی  $x = \{x_1, \dots, x_n\}$  فاصله دنباله  $x$  با نزدیک‌ترین بردار کتاب کد، محاسبه می‌شود. اگر  $i$  امین دنباله با  $x_i$  و نزدیک‌ترین بردار کتاب کد با  $y_i$  نشان داده شود، مربع فاصله اقلیدسی بین دو بردار به صورت زیر خواهد بود.

$$d(x_i, y_i) = \frac{1}{n} \sum_{i=1}^n (x_i - y_i)^2 \quad (11)$$

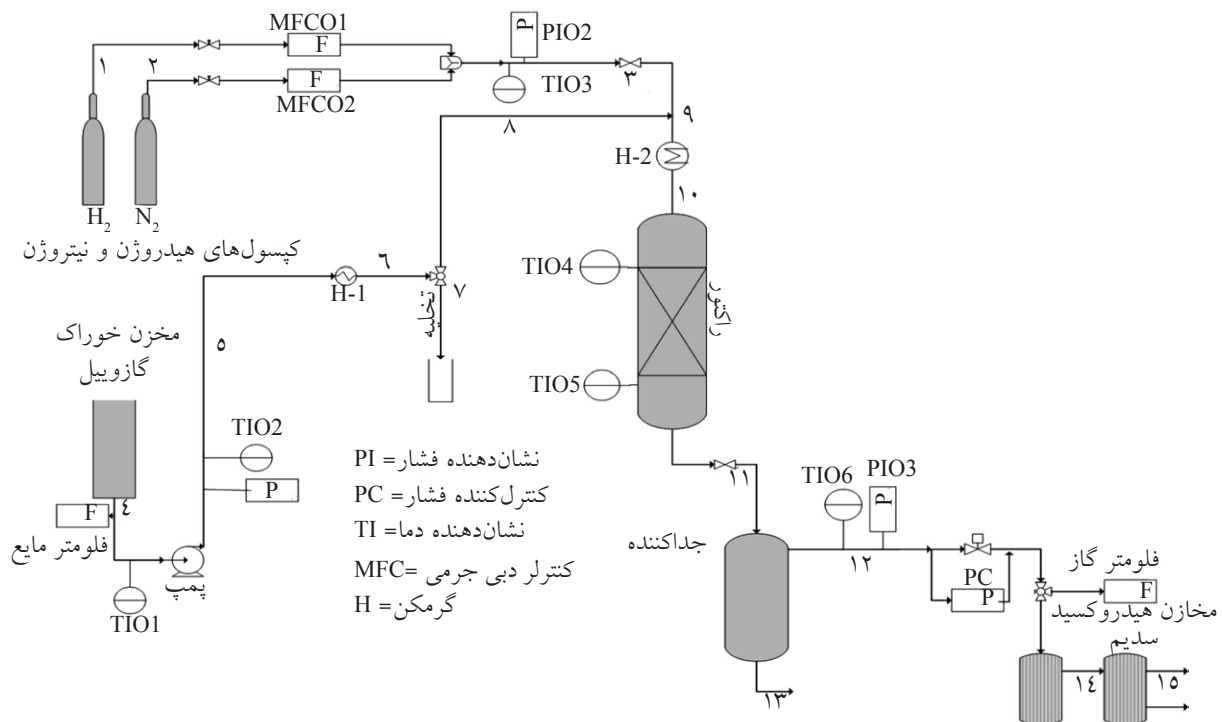
بنابراین خواهیم داشت:

$$S(x) = \sum_{i=1}^n \min_j d(x_i, y_j) \quad (12)$$

در صورت تعریف فاکتور وزنی، میانگین مربع خطای وزنی از رابطه ۱۱ به صورت زیر محاسبه خواهد شد:

$$m(x_i, y_i) = \frac{1}{n} \sum_{i=1}^n w_i (x_i - y_i)^2 \quad (13)$$

هر چه مقدار  $S(x)$  کمتر باشد، نمونه ورودی بهتر با مدل تطبیق یافته است. ترکیب روش کوانتیزاسیون برداری و ماشین بردار پشتیبان به منظور کاهش زمان آموزش و تست مورد استفاده قرار می‌گیرد [۲۷]. در این صورت به دلیل کاهش بردارهای آموزشی، محاسبات و در نتیجه زمان مورد نیاز برای تعیین بردارهای پشتیبان کاهش می‌یابد. همچنین استفاده از کوانتیزاسیون برداری، علاوه بر کاهش حجم داده‌های آموزشی، باعث چشم‌پوشی از داده‌هایی می‌شود که شباهت کمتری به سایر داده‌ها دارند. در نتیجه، می‌توان



مقدار	مشخصات
۷۲۰۰ (ppm)	مقدار گوگرد خوراک
۰/۰۱۲۷ m	قطر راکتور
۰/۶۳ m	طول راکتور
۰/۱۱ m	طول بستر کاتالیست

شکل ۲- شماتیک پایلوت گوگردزایی و مشخصات آن

دارند، به گونه‌ای که برای رسیدن به نتیجه دلخواه باید آنها را بسته به شرایط ورودی در محدوده مناسبی کنترل نمود. به برخی از این متغیرها در ادامه اشاره می‌گردد.

#### دمای راکتور

دمای ورودی راکتور معمولاً به سادگی توسط کاربر کنترل می‌شود و مقدار آن باید به گونه‌ای باشد که جداسازی اجزاء گوگردی به خوبی انجام گیرد. دمای خروجی راکتور تابعی از کیفیت خوراک بوده و به راحتی نمی‌توان آن را تغییر داد، مگر اینکه دمای ورودی تغییر نماید.

#### کیفیت و شدت جریان خوراک

میزان کاتالیست مورد نیاز مانند سایر پارامترهای طراحی به کمیت و کیفیت خوراک واحد بستگی دارد. به عنوان مثال با افزایش شدت جریان خوراک، لازم است دمای راکتور جهت حذف میزان مشابهی از سولفور افزایش یابد.

جهت کنترل دما و فشار سیستم توسط حسگرها، کنترلرهای نشانگر دما، نشانگرهای دما و همچنین کنترلرهای دبی جریان از طریق برنامه نرم‌افزاری به یک تابلوی الکترونیکی متصل گردیدند. این تابلو رابط بین نرم‌افزار جمع‌آوری اطلاعات و بخش‌های کنترلی سیستم می‌باشد. در پایلوت گوگردزایی یک مخزن خوراک متصل به یک بورت مدرج و دارای شیر تفلونی، یک پمپ دیافراگمی رفت و برگشتی و قابل تنظیم که در فشار بالا کار می‌کند، یک کندانسور دو جداره از فولاد ضد زنگ ۳۱۶ و دو مخزن حاوی هیدروکسید سدیم برای جذب سولفید هیدروژن از گازهای خروجی در نظر گرفته شده است. خطوط لوله و اتصالات نیز بر اساس استیل ۱/۴ اینچ انتخاب شده‌اند.

#### جمع‌آوری داده‌های سیستم

برخی از پارامترهای عملیاتی در فرآیند HDS اهمیت زیادی

سوخت دیزل اهمیت زیادی دارد. در صورتی که نسبت هیدروژن به گازوئیل از نسبت طراحی کمتر باشد، سرعت غیرفعال شدن کاتالیست افزایش می‌یابد. با توجه به توضیحات بالا، طراحی آزمایشات برای آزمایشات پایلوتی بر اساس محدوده های ذکر شده در جدول ۱ می‌باشد.

#### توسعه مازول حذف خطا و پیاده‌سازی الگوریتم هیبرید

شکل ۳ الگوریتم روش ترکیبی پیش‌پردازش داده‌ها و ماشین‌بردار پشتیبان که پارامترهای آن با استفاده از روش بهینه‌سازی جستجوی شبکه‌ای<sup>۱</sup> تنظیم گردیده را نشان می‌دهد. تنظیم فرا پارامترهای مدل در ماشین‌بردار رگرسیون، در پیش‌بینی نتایج خروجی نیز تأثیرگذار می‌باشد. در اینجا برای بهینه‌سازی این پارامترها از روش‌های بهینه‌سازی جستجوی شبکه‌ای استفاده شده است.

همچنین با افزایش شدت جریان خوراک به منظور ثابت نگه‌داشتن نسبت هیدروژن به گازوئیل، باید میزان جریان گاز افزایش یابد. افزایش دما باعث شدت یافتن واکنش‌های تشکیل کک و در نتیجه کاهش زمان عمر کاتالیست می‌گردد. از طرف دیگر در صورت کمتر بودن شدت جریان خوراک نسبت به ظرفیت طراحی، به دلیل اینکه خوراک مدت زمان بیشتری با کاتالیست در تماس است، میزان تشکیل کک افزایش می‌یابد.

#### فشار راکتور

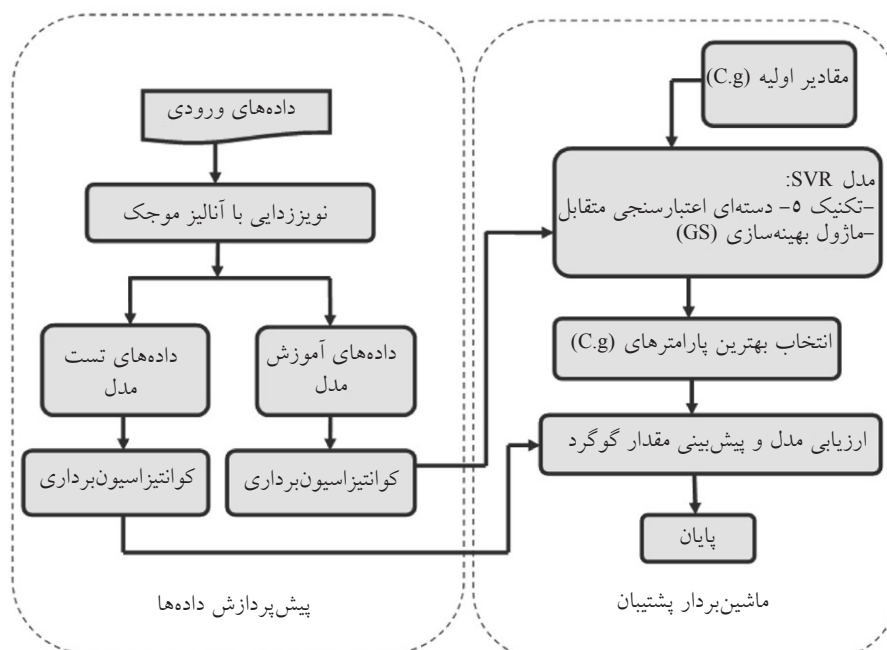
میزان فشار در واکنش‌های فرآیند گوگردزدایی از اهمیت بالایی برخوردار است. کاهش فشار راکتور به زیر میزان طراحی، موجب کاهش فعالیت کاتالیست و همچنین افزایش سرعت غیرفعال شدن کاتالیست به دلیل تشکیل کک می‌گردد.

#### نسبت هیدروژن به گازوئیل ( $H_2/Gasoil$ )

نسبت هیدروژن به گازوئیل در فرآیند گوگردزدایی از

جدول ۱- سطوح پارامترها

سطح ۵	سطح ۴	سطح ۳	سطح ۲	سطح ۱	پارامتر
	۳۷۰	۳۵۰	۳۳۵	۳۲۰	دمای ورودی به راکتور ( $^{\circ}C$ )
		۷۰	۶۰	۵۰	فشار راکتور (bar)
۱۷۰	۱۴۰	۱۲۰	۱۰۰	۸۵	نسبت $H_2/Gasoil$ ( $nm^3/m^3$ )
۰/۳۲	۰/۲۹	۰/۲۶	۰/۲۳	۰/۲۰	دبی جریان مایع (cc/min)



شکل ۳- الگوریتم بلوک‌های مدل (روش ترکیبی پیش‌پردازش داده‌ها و ماشین‌بردار پشتیبان)

## مدل حسگر مجازی

برای ساخت مدل حسگر مجازی از متغیرهای ورودی دبی هیدروژن، دبی گازوییل، فشار راکتور، دمای ورودی به راکتور و دمای خروجی از راکتور و متغیر خروجی غلظت گوگرد خروجی از راکتور استفاده گردید. در اینجا پس از رسیدن فرآیند به حالت پایا، داده‌برداری از گوگرد خروجی در فواصل ۴ ساعته انجام می‌شود. سپس بر اساس طراحی آزمایشات صورت گرفته و تغییر یکی از متغیرهای عملیاتی با شیب مناسب به شرایط جدید می‌رسیم. در حالت تغییر شرایط عملیاتی، از محصول خروجی نمونه‌گیری نمی‌شود. بنابراین کندانسور تخلیه شده و بعد از رسیدن فرآیند به حالت پایای جدید، با گذشت زمان ۴ ساعت، داده بعدی اخذ می‌شود. لازم به ذکر است داده‌های مربوط به دما، فشار و دبی در فواصل یک دقیقه از طریق سیستم داده برداری پایلوت بر روی دیسک سخت ذخیره‌سازی می‌گردد که با تکنیک آنالیز موجک نویزهای آنها پالایش می‌شود. در این تحقیق برای ساخت مدل ماشین‌بردار رگرسیون از نرم‌افزار *LIBSVM* استفاده گردید [۲۸].

نرم‌افزار *LIBSVM* یک نرم‌افزار یکپارچه است که برای حالت‌های ماشین‌بردار طبقه‌بندی<sup>۱</sup>، رگرسیون و برآورد توزیع<sup>۲</sup> به کار می‌رود. این نرم‌افزار برای حالت ماشین‌بردار رگرسیون دو حالت *v-SVR* و *ε-SVR* را شامل می‌شود. در این تحقیق پس از پیاده‌سازی پیش‌پردازش داده‌ها با استفاده از روش‌های مختلف آنالیز موجک و انتخاب بهترین روش در ایجاد داده‌های مطمئن و معتبر، مراحل اجرای مدل و انجام محاسبات با استفاده از حالت *ε-SVR* توسط نرم‌افزار *LIBSVM* انجام پذیرفت. لازم به ذکر است در این تحقیق کلیه محاسبات در محیط نرم‌افزاری *MATLAB 7.8* و لینک آن به نرم‌افزار فوق‌الذکر صورت گرفت. در اینجا از هیبرید روش کوانتیزاسیون برداری و ماشین‌بردار پشتیبان به منظور کاهش زمان آموزش و تست و افزایش دقت پیش‌بینی استفاده گردید. همچنین برای بهینه‌سازی پارامترهای مدل از روش بهینه‌سازی جستجوی نقطه‌ای بهره گرفته شد.

مراحل اصلی توسعه مدل به قرار زیر است:

- ورود داده‌های آموزشی و تست

- به کارگیری تکنیک کوانتیزاسیون برداری

- به کارگیری تکنیک اعتبارسنجی - متقابل<sup>۳</sup> (در این مطالعه از تکنیک ۵-دسته‌ای اعتبارسنجی - متقابل<sup>۴</sup> استفاده گردیده است)

- انتخاب نوع ماشین‌بردار پشتیبان (در این مطالعه از مدل *ε-SVR* استفاده شده است)

- انتخاب نوع هسته کرنل (در این مطالعه از هسته کرنل گوسی استفاده گردیده است)

- بهینه‌سازی پارامترهای مدل  $g$  و  $C$  با استفاده از روش بهینه‌سازی نقطه‌ای

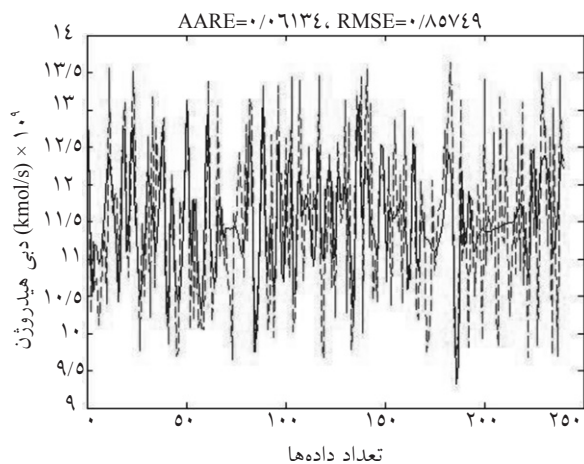
- معتبرسازی مدل و پیش‌بینی نتایج

## نتایج و بحث

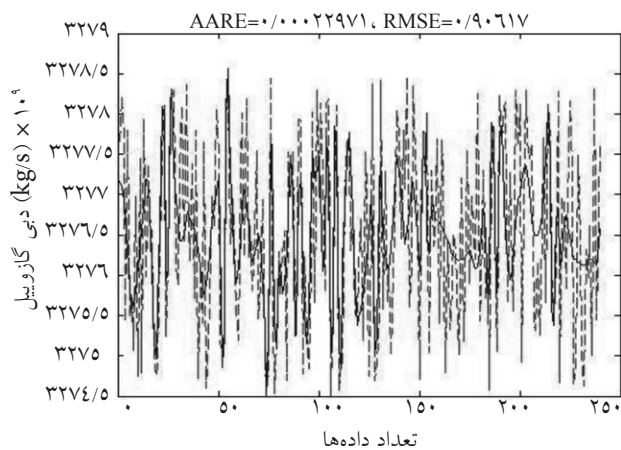
در این تحقیق سعی شده با استفاده از خانواده‌های مختلف موجک‌ها نویزهای احتمالی موجود در داده‌ها حذف شود. بدین منظور روش‌های مختلف حذف نویز با استفاده از چند خانواده معروف از موجک‌ها مورد استفاده قرار گرفته است و نتایج حاصل از عملیات حذف نویز از داده‌های سیستم با یکدیگر مقایسه شده است. آزمایشات با تابع پایه‌ها و دابچیز و با الگوریتم‌های انتخاب آستانه *Minimaxi*، *Sqtwolog*، *Heursure*، *Rigrsure* انجام شده است. نمونه‌ای از نتایج در شکل‌های ۴ و ۵ و جدول ۲ ارائه شده است. مقایسه نتایج نشان می‌دهد که روش *Db4* به همراه روش آستانه‌گیری *Rigrsure* بهترین نتایج حذف خطا را به دنبال دارد. همان‌گونه که پیشتر عنوان شد، در این مطالعه از سیستم ترکیبی متشکل از کوانتیزاسیون برداری و ماشین‌بردار پشتیبان جهت کاهش زمان آموزش و آزمایش و همچنین نرمال کردن داده‌ها استفاده شده است. در شکل ۶ پیش‌بینی ماشین‌بردار رگرسیون در دو حالت استفاده از کوانتیزاسیون برداری و بدون آن رسم شده است. همان‌طور که در این شکل مشخص است، این خطاها در دقت پیش‌بینی مدل تأثیر بسزایی خواهد داشت.

1. Support Vector Classification (SVC)
2. Support Vector Regression (SVR)
3. Cross-Validation
4. 5-fold Cross-Validation

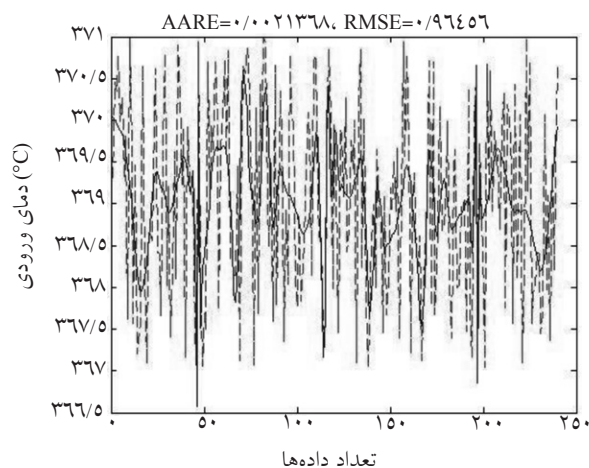




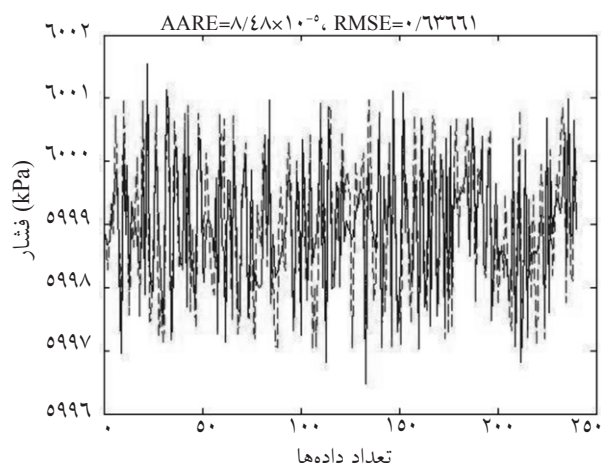
ب) پیاده‌سازی آنالیز موجک روی دبی هیدروژن



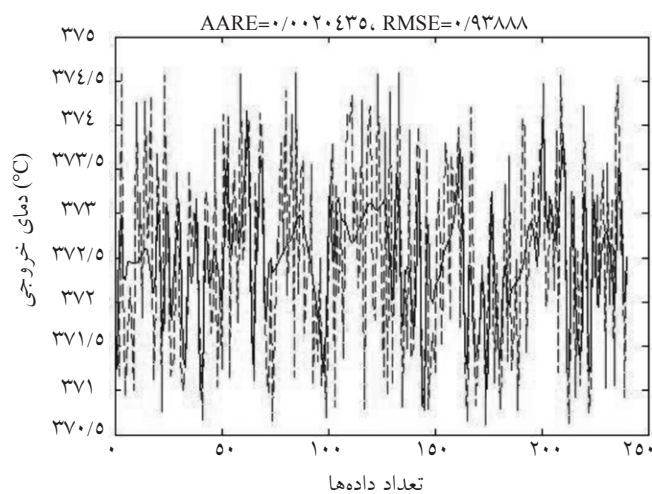
الف) پیاده‌سازی آنالیز موجک روی دبی گازوییل



د) پیاده‌سازی آنالیز موجک روی دمای ورودی راکتور

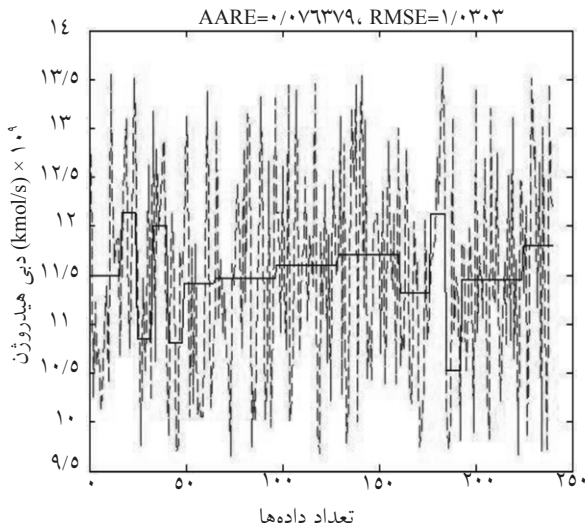


ج) پیاده‌سازی آنالیز موجک روی فشار راکتور

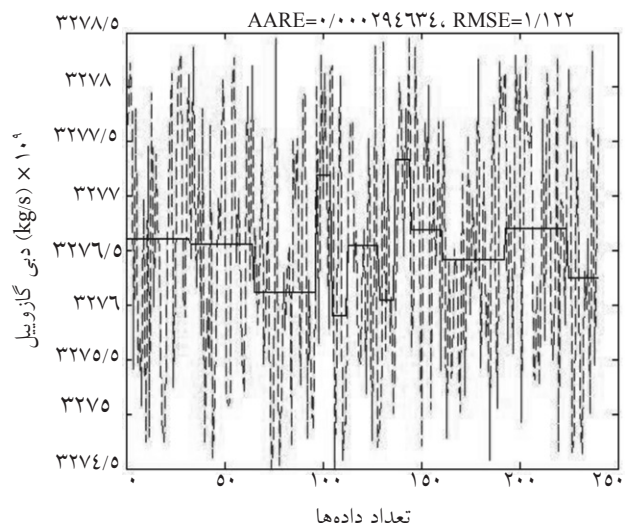


ه) پیاده‌سازی آنالیز موجک روی دمای خروجی راکتور

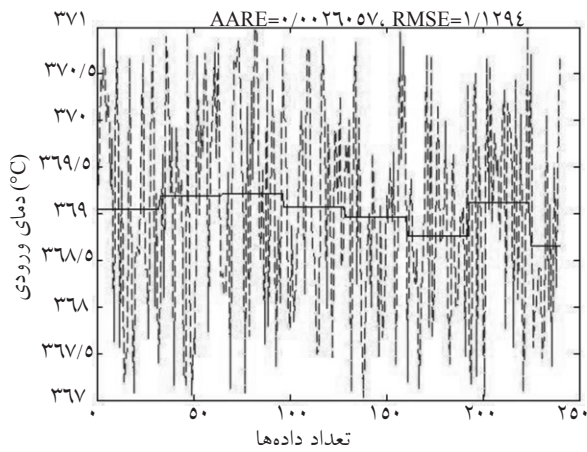
شکل ۴- نمونه‌ای از پیاده‌سازی آنالیز موجک بر روی ۲۴۰ داده در زمان ۴ ساعت (Db4(one)-Rigsure) (خطوط نقطه‌چین، سیگنال نویزدار و خطوط پیوسته، سیگنال بدون نویز)



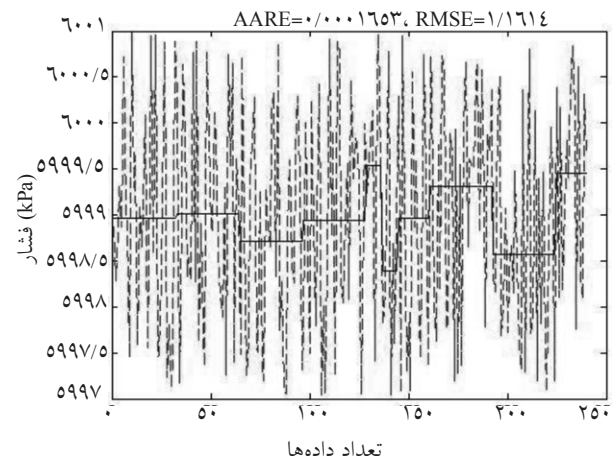
(ب) پیاده‌سازی آنالیز موجک روی دبی هیدروژن



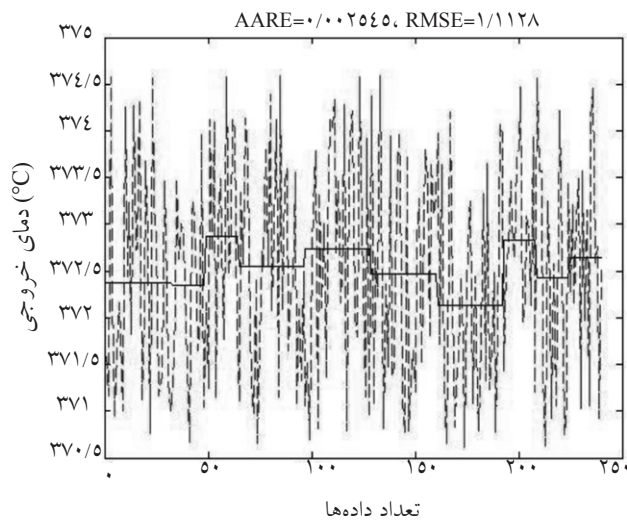
(الف) پیاده‌سازی آنالیز موجک روی دبی گازوئیل



(د) پیاده‌سازی آنالیز موجک روی دمای ورودی راکتور



(ج) پیاده‌سازی آنالیز موجک روی فشار راکتور

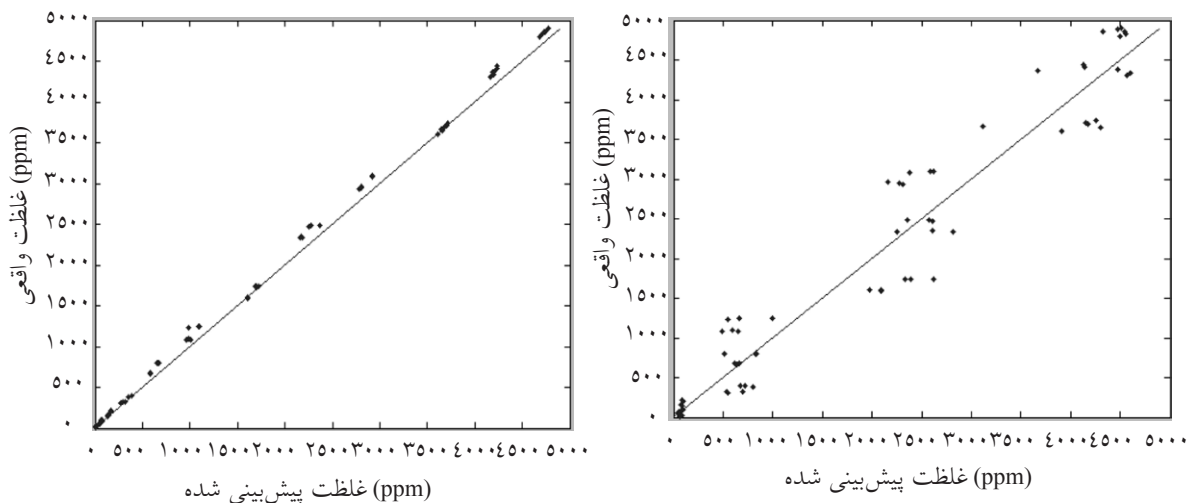


(ه) پیاده‌سازی آنالیز موجک روی دمای خروجی راکتور

شکل ۵- نمونه‌ای از پیاده‌سازی آنالیز موجک بر روی ۲۴۰ داده در زمان ۴ ساعت (Harr(mln)-Heursure) (خطوط نقطه‌چین، سیگنال نویزدار و خطوط پیوسته، سیگنال بدون نویز)

جدول ۲- مقایسه روش‌های مختلف در آنالیز موجک

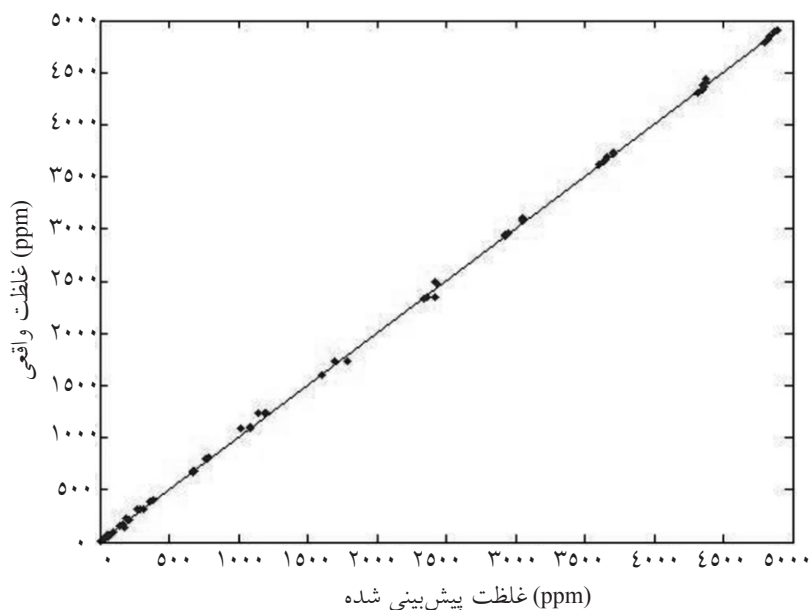
ردیف	نوع آنالیز موجک	روش آستانه‌گیری	متغیر	AARE	RMSE
۱	Db4(mln)	Heursure	$Q_{gasoil} * 1E9(kg/s)$	۰/۰۰۰۲۹۰	۱/۱۱۰۰
			$QH_2 * 1E9(kmol/s)$	۰/۰۷۷۸	۱/۰۵۵۸
			Pressure(kPa)	۰/۰۰۰۱۶۴	۱/۱۵۱۵
			$T_{in}(^{\circ}C)$	۰/۰۰۲۴	۱/۰۷۴۰
			$T_{out}(^{\circ}C)$	۰/۰۰۲۵	۱/۱۰۹۱
۲	Harr(mln)	Heursure	$Q_{gasoil} * 1E9(kg/s)$	۰/۰۰۰۲۹۴	۱/۱۲۲۰
			$QH_2 * 1E9(kmol/s)$	۰/۰۷۶۴	۱/۰۳۰۳
			Pressure(kPa)	۰/۰۰۰۱۶۵	۱/۱۶۱۴
			$T_{in}(^{\circ}C)$	۰/۰۰۲۶	۱/۱۲۹۴
			$T_{out}(^{\circ}C)$	۰/۰۰۲۵	۱/۱۱۲۸
۳	Harr	Minimaxi	$Q_{gasoil} * 1E9(kg/s)$	۰/۰۰۲۴۷	۰/۹۷۴۴
			$QH_2 * 1E9(kmol/s)$	۰/۰۶۳۵	۰/۸۸۳۳
			Pressure(kPa)	۰/۰۰۰۱۵۸	۱/۱۱۷۵
			$T_{in}(^{\circ}C)$	۰/۰۰۲۴	۱/۰۳۵۹
			$T_{out}(^{\circ}C)$	۰/۰۰۲۴	۱/۰۴۲۷
۵	Db4(one)	Rigrsure	$Q_{gasoil} * 1E9(kg/s)$	۰/۰۰۰۲۲۹	۰/۹۰۶۲
			$QH_2 * 1E9(kmol/s)$	۰/۰۶۱۳	۰/۸۵۷۹
			Pressure(kPa)	۰/۰۰۰۸۴۸	۰/۶۳۶۶
			$T_{in}(^{\circ}C)$	۰/۰۰۲۱	۰/۹۶۴۶
			$T_{out}(^{\circ}C)$	۰/۰۰۲۰	۰/۹۳۸۹
۶	Harr(mln)	Sqtwolog	$Q_{gasoil} * 1E9(kg/s)$	۰/۰۰۰۳۰۴	۱/۱۴۶۵
			$QH_2 * 1E9(kmol/s)$	۰/۰۷۷۶	۱/۰۵۴۸
			Pressure(kPa)	۰/۰۰۰۱۶۷	۱/۱۷۰۶
			$T_{in}(^{\circ}C)$	۰/۰۰۲۶	۱/۱۲۹۴
			$T_{out}(^{\circ}C)$	۰/۰۰۲۵	۱/۱۱۹۲



شکل ۶- مقایسه پیش‌بینی ماشین‌بردار رگرسیون در حالت استفاده از الف) کوانتیزاسیون برداری و ب) بدون استفاده از کوانتیزاسیون

شده است. در این روش لازم است که مقادیر  $C$  و  $g (1/2\sigma^2)$  بهینه گردند برای این منظور مقدار  $\log_2 C \in [-5, 20]$  و مقدار  $\log_2 g \in [-30, 20]$  انتخاب گردید و در نهایت مقدار  $C$  و  $g$  به ترتیب برابر ۱۲۸ و ۰/۰۶۲۵ به دست آمد. همچنین مقدار  $AARE$  برابر ۰/۰۵۳ می‌باشد که نشان از دقت بالای مدل پیش‌بین می‌باشد. همان‌گونه که در شکل مشخص است، در یک محدوده وسیع می‌توان غلظت گوگرد را به خوبی پیش‌بینی نمود.

در این مطالعه پس از معتبرسازی داده‌ها، آنها به دو قسمت آموزش و تست تقسیم شدند. لازم به توضیح است که از مجموع ۳۰۰ داده، تعداد ۲۳۰ داده برای آموزش و ۷۰ داده برای تست استفاده شد. در ادامه لازم است پارامترهای مدل ماشین‌بردار بهینه گردد. بنابراین در این جا از روش الگوریتم جستجوی نقطه‌ای جهت بهینه‌سازی مقادیر پارامترهای مدل بهره گرفته شد. در شکل ۷ منحنی دقت<sup>۱</sup> مربوط به مدل ماشین‌بردار پشتیبان که با الگوریتم جستجوی نقطه‌ای پارامترهای آن تنظیم گردیده، نشان داده



شکل ۷- منحنی‌های دقت مدل پیش‌بین

## نتیجه‌گیری

پشتیبان در تعیین غلظت گوگرد خروجی از راکتور تصفیه هیدروژنی استفاده کرد و فرآیند را تا حد نهایی عملکرد مورد بهره‌برداری قرار داد.

## علائم و نشانه‌ها

- $K(x,y)$ : تابع کرنل
- $\psi_{a,b}(t)$ : تابع موجک مادر
- $T_{out}$ : دمای خروجی راکتور
- $T_{in}$ : دمای ورودی راکتور
- $Q_{H_2}$ : دبی هیدروژن
- $Q_{gasoil}$ : دبی گازوئیل
- GS: روش جستجوی نقطه‌ای
- HDS: فرآیند تصفیه هیدروژنی
- VQ: کوانتیزاسیون برداری
- SVM: ماشین بردار پشتیبان
- SVR: ماشین بردار رگرسیون
- SVC: ماشین بردار دسته‌بندی
- $\xi_1^+$ ,  $\xi_1^-$ : متغیرهای کمبود
- C: پارامتر تنظیم‌کننده
- $\sigma^2$ : پارامتر تابع کرنل
- $g(1/2\sigma^2)$ : پارامتر تابع کرنل

ضرورت تولید محصول با کیفیت هدف اساسی صنایع پالایشی می‌باشد. از طرفی آنالایزر و حسگرهای سخت‌افزاری که در صنعت برای این منظور به کار می‌روند، بسیار هزینه بر بوده و زود خراب می‌شوند. برای حل این مشکل می‌توان از حسگرهای مجازی بهره برد و استفاده از این نوع حسگرهای نرم‌افزاری که بر اساس داده‌های صحیح و معتبر ورودی کار می‌کند، می‌تواند در این خصوص کمک فراوانی به مهندسان بهره‌بردار نماید. در این پژوهش به منظور پیش‌بینی غلظت گوگرد، یکی از روش‌های جدید یادگیری ماشین به نام ماشین بردار رگرسیون به کار گرفته شد.

لازم به ذکر است که برای آموزش و تست ماشین بردار رگرسیون لازم است که از داده‌های صحیح استفاده شود. بنابراین در این تحقیق برای حذف خطاهای کوچک از روش آنالیز موجک استفاده شد و برای تعیین نوع آنالیز موجک از انواع مختلف این آنالیز استفاده گردید و میزان خطای آنها با یکدیگر مقایسه شد. همچنین از کوانتیزاسیون برداری جهت کاهش زمان آموزش و آزمایش و نرمال کردن داده‌ها استفاده گردید. معیار عملکرد AARE در نتایج خروجی دقت بالای مدل را نشان می‌دهد. بنابراین از این روش هیبرید می‌توان در پیش‌بینی دقیق‌تر ماشین بردار

## مراجع

- [1]. Kadlec P., Gabrys B. and Strandt S., "Data-driven soft sensors in the process industry", *Comput. Chem. Eng.*, Vol. 33, pp. 795, 2009.
- [2]. Pan T. H., Wong D. S. H. and Jang S. S., "Development of a novel soft sensor using a local model network with an adaptive subtractive clustering approach", *Ind. Eng. Chem. Res.*, Vol. 49, pp. 4738-4747, 2010.
- [3]. Gonzagaa J. C. B., Meleirob L. A. C., Kianga C. and Filho R. M., "ANN-based soft-sensor for real-time process monitoring and control of an industrial polymerization process", *Comput. Chem. Eng.* Vol. 33, pp. 43-49, 2009.
- [4]. Chitralakha S. B. and Shah S.L., "Application of support vector regression for developing soft sensors for non-linear processes", *The Canadian Journal of Chemical Engineering*. Vol. 88, pp. 696-709, 2010.
- [5]. Park T. C., Kim T. Y., and Yeo Y. K., "Prediction of the melt flow index using partial least squares and support vector regression in high-density polyethylene (HDPE) process", *Korean J. Chem. Eng.*, 27(6), pp. 1662-1668, 2010.
- [6]. Dutta S. and Gupta J. P., "PVT correlations of indian crude using support vector regression", *Energy & Fuels*,

Vol. 23, pp. 5483–5490, 2009.

[7]. Yan W., Shao H. and Wang X., "Soft sensing modeling based on support vector machine and Bayesian model selection", *Comput. Chem. Eng.*, Vol. 28, pp. 1489–1498, 2004.

[8]. Desai K., Badhe Y., Tambe S. S. and Kulkarni B. D., "Soft-sensor development for fed-batch bioreactors using support vector regression", *Biochemical Engineering Journal* 27, pp. 225–239, 2006.

[9]. Guohai L., Dawei Z., Haixia X. and Congli M., "Model optimization of SVM for a fermentation soft sensor", *Expert Systems with Applications*, Vol. 37, pp. 2708–2713, 2010.

[10]. Liu Y., Hu N., Wang H. and Li P., "Soft chemical analyzer development using adaptive least-squares support vector regression with selective pruning and variable moving window size", *Ind. Eng. Chem. Res.*, Vol. 48, pp. 5731–574, 2009.

[11]. Hong W. C., "Traffic flow forecasting by seasonal SVR with chaotic simulated annealing algorithm", *Neurocomputing*, Vol. 74, 2096-2107, 2011.

[12]. Minqiang P., Dehuai Z. and Gang X., "Temperature prediction of hydrogen producing reactor using SVM regression with PSO", *Journal of computers*, Vol 5, No.3, 2010.

[13]. Yin J., "LogP prediction for blocked tripeptides with amino acids descriptors (HMLP) by multiple linear regression and support vector regression", *Procedia Environmental Sciences* 8, pp. 173–178, 2011.

[14]. Vapnik V. N. *The nature of statistical learning theory*, 2<sup>nd</sup>, ed. New York: Springer; 1999.

[15]. Boser B. E., Guyon I. M. and Vapnik V. N. *A training algorithm for optimal margin classifiers*, In D. Haussler, editor, 5<sup>th</sup> Annual ACM Workshop on COLT, 144-152, Pittsburgh, PA, ACM Press, 1992.

[16]. Vapnik V. N., *Statistical learning theory*, Wiley, New York, 1998.

[17]. Basak D., Pal S. and Patranabis D. C., "Support vector regression", *Neural Inf. Process.* Vol. 11, pp. 203–225, 2007.

[18]. Cherkassky V. and Ma Y., "Practical selection of SVM parameters and noise estimation for SVM regression", *Neural Networks*, Vol. 17, pp. 113–126, 2004.

[19]. Benqlilou C., *Data reconciliation as a framework for chemical processes optimization and control*, PHD Thesis, March, 2004.

[20]. Singh M. K., *Denising of natural images using the wavelet transform*, Master's Thesis. San Jose State University, 2010.

[21]. Huang H. P. and Luo K. Y., "On-line wavelets filtering with application to linear dynamic data reconciliation", *Ind. Eng. Chem. Res.* Vol. 46, pp. 8746-8755, 2007.

[22]. Unser M. and Blu T., "Wavelet theory demystified", *IEEE Transaction on Signal Processing*, 51(2), pp. 470-483, 2003.

[23]. Shukla P. D., *Complex wavelet transforms and their applications*, Master of Philosophy Thesis, University of Strathclyde, 2003.

[24]. Phinyomark A., Limsakul C., and Phukpattaranont P., "A comparative study of wavelet denoising for multi-function myoelectric control", in *International Conference on Computer and Automation Engineering*, pp. 21–25,

2009.

[25]. Jiang C. F. and Kuo S. L., "A comparative study of wavelet denoising of surface electromyographic signals", in 29<sup>th</sup> Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 1868–1871, 2007.

[26]. Somasundaram K. and Vimala S., "Fast encoding algorithm for vector quantization", International Journal of Engineering Science and Technology, Vol. 2(9), pp. 4876-4879, 2010.

[27]. Yu T., Simoff S. and Jan T., "VQSVM: a case study for incorporating prior domain knowledge in to inductive machine learning", Neurocomputing Vol. 73, pp. 2614–2623, 2010.

[28]. Chang C. C. and Lin C. J., *LIBSVM: a library for support vector machines*, Software <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>, 2001.