

کاربرد الگوریتم ماشین بردار پشتیبان و تبدیل موجک برای مدل‌سازی فرآیند بارش - رواناب (مطالعه موردی: آقچای - ایران)

مهدی کماسی^{۱*} و سروش شرقی^۲

چکیده

اهمیت مدیریت منابع آب، نیاز به مدل‌سازی دقیق فرآیند بارش و رواناب را در دهه گذشته مطرح کرده‌است. در این پژوهش برای مدل‌سازی فرآیند بارش و رواناب از ترکیب مدل ماشین بردار پشتیبان و تبدیل موجک بهره گرفته شده است. بدین‌منظور سری‌های زمانی بارش و رواناب با آنالیز موجک به چندین زیرسری با مقیاس‌های زمانی مختلف تبدیل شده و این زیرسری‌های زمانی به عنوان ورودی مدل ماشین بردار پشتیبان برای پیش‌بینی رواناب روزانه در نظر گرفته شده است. نتایج حاصل از صحت‌سنجی مدل‌ها بیانگر آن است که بیشترین مقدار ضریب تبیین (R^2) و کمترین مقدار جذر میانگین مربع خطا (RMSE) برای مدل منفرد ماشین بردار پشتیبان به ترتیب ۰/۶۹ و ۰/۷۹ و برای مدل ترکیبی ماشین بردار پشتیبان موجکی به ترتیب ۰/۹۲ و ۰/۳۵ است. دلیل برتری مدل ترکیبی نسبت به مدل منفرد ناشی از این است که مدل ترکیبی ماشین بردار پشتیبان موجکی، به جای استفاده از سری زمانی داده‌های بارش و رواناب در یک مقیاس کلی، از چندین زیرسری پردازش شده زمانی با درجات تجزیه مختلف، به عنوان ورودی در مدل استفاده می‌کند. همچنین نتایج نشان داد که مدل ترکیبی ماشین بردار پشتیبان موجکی در مقایسه با سایر مدل‌های ترکیبی مانند شبکه عصبی مصنوعی موجکی (WANN) دارای کارایی و دقت بالاتری است.

واژه‌های کلیدی: تبدیل موجک، تابع موجک، چند مقیاس‌ای، خودهمبستگی، درجه تجزیه، ماشین بردار پشتیبان.

ارجاع: کماسی م. و سروش شرقی س. ۱۳۹۶. کاربرد الگوریتم ماشین بردار پشتیبان و تبدیل موجک برای مدل‌سازی فرآیند بارش - رواناب (مطالعه موردی: آقچای - ایران). مجله پژوهش آب ایران. ۲۵: ۵۱-۶۰.

1- استادیار گروه مهندسی عمران، دانشکده فنی و مهندسی، دانشگاه آیتاله العظمی بروجردی (ره)، بروجرد.

2- دانشجوی کارشناسی ارشد مهندسی عمران، سازه‌های هیدرولیکی، دانشکده فنی و مهندسی، دانشگاه آیتاله العظمی بروجردی (ره).

* نویسنده مسئول: komasi@abru.ac.ir

تاریخ پذیرش: ۱۳۹۴/۰۸/۲۴

تاریخ دریافت: ۱۳۹۴/۱۱/۱۴

مقدمه

می‌شود (ونگ، ۲۰۰۵) اما گاهی طبقه‌بندی و رگرسیون داده‌ها به صورت خطی ممکن نیست. بنابراین مدل SVM از ابزاری به نام تابع کرنل یا هسته برای تبدیل و انتقال داده‌ها به فضایی با ویژگی ابعادی بالاتر استفاده می‌کند به طوری که جداسازی خطی داده‌ها ممکن باشد (برجس، ۱۹۹۹). هرچند مدل‌های هوش مصنوعی همانند SVM کارایی مناسبی را در مدل‌سازی سری‌های زمانی هیدرولوژیکی ایستا دارد (ژانگ و دانگ، ۲۰۰۱)، اما این روش‌ها در مدل‌سازی داده‌های ناپایستا که گستره وسیعی از مقیاس‌های زمانی را از سال و ماه گرفته تا روز و ساعت دربر می‌گیرند، دارای دقت مناسبی نمی‌باشند. در چنین مواقعی اگر روی داده‌های ورودی عمل پیش‌پردازش انجام نشود و هرکدام از مقیاس‌های زمانی یاد شده جداگانه تفکیک نشود، مدل SVM توانایی درک رفتار و روند حاکم بر سری ناپایستا را نخواهد داشت. بنابراین پیش‌پردازش و تفکیک داده‌های زمانی به مقیاس‌های زمانی مختلف می‌تواند درک درست‌تری از الگوهای موجود در داده‌های زمانی را ارائه دهد (رجایی و همکاران، ۲۰۱۰؛ نورانی و همکاران، ۲۰۱۲). از این‌رو در این پژوهش از مدل ترکیبی تبدیل موجک به عنوان پیش‌پردازش‌کننده داده‌های سری زمانی و مدل SVM برای مدل‌سازی فرآیند بارش - رواناب استفاده شده است. در ادامه این مقاله، در بخش دوم مفاهیم مربوط به SVM و تابع موجک مرور و سپس ساختار مدل ترکیبی ماشین بردار پشتیبان موجکی (WSVM)^۶ ارائه شده است. در بخش سوم، حوضه مورد بررسی معرفی شده و در بخش چهارم روی کارایی‌های مدل منفرد SVM و مدل ترکیبی WSVM بحث و بررسی و در نهایت نتیجه‌گیری و جمع‌بندی در قسمت پایانی این مقاله ارائه شده است.

مواد و روش‌ها

ماشین بردار پشتیبان (SVM)

ماشین‌های بردار پشتیبان همانند دیگر روش‌های هوش مصنوعی بر مبنای الگوریتم داده‌کاوی عمل می‌کنند. مهم‌ترین وظایف مدل ماشین‌های بردار پشتیبان، طبقه‌بندی و خطی‌سازی یا رگرسیون داده‌ها است. خطی‌سازی داده‌ها با الگوریتم خطی‌سازی ε -SVR و با معرفی تابع ε -sensitive انجام می‌شود. در یک الگوریتم

شبیه‌سازی دقیق و واقعی از فرآیندهای هیدرولوژیکی مانند بارش و رواناب می‌تواند اطلاعات مؤثری را در ارتباط با برنامه‌ریزی‌های شهری، بهره‌وری از زمین و مدیریت منابع آب برای یک حوضه فراهم کند. بنابراین مدل‌های هیدرولوژیکی زیادی برای شبیه‌سازی این فرآیند پیچیده تکامل یافته است. یکی از مدل‌های کلاسیک برای مدل‌سازی فرآیندهای هیدرولوژیکی بهره‌گیری از مدل خطی آریما (ARIMA)^۱ است (سالاس و همکاران، ۱۹۸۰). این مدل توانایی رصد نمودن داده‌های ایستا را دارد اما توانایی دریافت الگوی غیرخطی پدیده‌های هیدرولوژیکی را ندارد. بنابراین در نسل بعدی، روش مدل‌سازی داده‌ها بر پایه نگرش مدل‌های جعبه سیاه یا هوش مصنوعی (AI)^۲، مورد توجه پژوهش‌گران قرار گرفت. یکی از شاخص‌ترین مدل‌ها در این نگرش، مدل شبکه عصبی مصنوعی (ANN)^۳ بوده که دارای کاربردهای زیادی در زمینه‌های مختلف پیش‌بینی فرآیندهای هیدرولوژیکی است (راجورکار و همکاران، ۲۰۰۴؛ گاربرجت، ۲۰۰۶). نسل بعدی از روش‌های AI، ماشین‌بردار پشتیبان (SVM)^۴ نام دارد که نسبت به روش ANN سریع‌تر آموزش دیده و نتایج دقیق‌تری را ارائه می‌کند (لین و همکاران، ۲۰۰۹). بنابراین امروزه روش SVM به عنوان ابزاری نوین برای مدل‌سازی سری‌های زمانی مطرح شده است. این مدل برای اولین بار از سوی وپنیک و کورتس در سال ۱۹۹۵ بر پایه تئوری یادگیری آماری ارائه شد. پژوهش‌های انجام شده گویای این واقعیت است که روش SVM سریع‌تر از ANN و سیستم استنتاج فازی عصبی- تطبیقی (ANFIS)^۵ آموزش می‌بیند. یکی دیگر از مزایای روش SVM در مقایسه با سایر روش‌ها این است که این مدل نیاز به داده‌های آموزشی کمتری برای مدل‌سازی دارد. همچنین در روش SVM زمان اجرا نسبت به سایر روش‌ها کوتاه‌تر است (چوی و چان، ۲۰۰۳). در روش SVM یکسری بردار رگرسیونی که بر پایه تئوری یادگیری آماری و حداقل کردن خطاهای طبقه‌بندی داده‌های آموزشی (آبه، ۲۰۰۵) استوار است، روی داده‌ها ایجاد

1- Auto Regressive Integrated Moving Average

2- Artificial Intelligence

3- Artificial Neural Network

4- Support Vector Machine

5- Adaptive Neural-Fuzzy Inference System

6- Wavelet-Support Vector Machine

در رابطه (۳)، C که خطای تولید نام دارد، در حقیقت رابطه جایگزین را برای تابع $f(x)$ مشخص کرده و مقدارش تا مقدار مجاز انحراف ε افزایش می‌یابد. در نهایت با استفاده از توابع کرنل $k(x_i, x)$ و پارامترهای لاگرانژین λ_i و λ_i' رابطه تابع خطی‌سازی، به شکل رابطه (۴) بازنویسی خواهد شد.

$$f(x, b) = \sum_{i=1}^k (I_i' - I_i) k(x_i, x) + b \quad (4)$$

بیان این نکته ضروری است که قبل از خطی‌سازی داده‌ها یا همان تقسیم خطی، برای اینکه ماشین بردار پشتیبان بتواند داده‌های با پیچیدگی بالا را پیش‌بینی کند، داده‌ها با تابع Φ به فضای با ابعاد خیلی بالاتر منتقل می‌شوند (شکل ۱). در این راستا برای حل مسأله با ابعاد خیلی بالا از قضیه دوگانگی لاگرانژ برای تبدیل مسأله کمینه‌سازی مورد نظر به فرم دوگانگی آن استفاده می‌شود که در آن به جای تابع پیچیده Φ که به فضایی با ابعاد بالا می‌رود، تابع ساده‌تری به نام تابع هسته (کرنل)، که ضرب برداری تابع Φ است، ظاهر می‌شود (رابطه (۴)). بنابراین کافی است در مسائل غیرخطی از کرنل مقدار ورودی به جای خود تابع استفاده شود.

خطی‌سازی ε -SVR، با فرض مقدار ورودی و خروجی x_i و y_i ، هدف یافتن تابعی است که کمترین انحراف ε را از y_i ها داشته باشد. اگر داده‌های آموزشی به تعداد K سری به صورت $(x_1, y_1), \dots, (x_k, y_k)$ مورد نظر باشند، تابع خطی‌سازی عبارت است از:

$$f(x) = (w, x) + b \quad \text{with } (w, x) \in R^N, b \in R \quad (1)$$

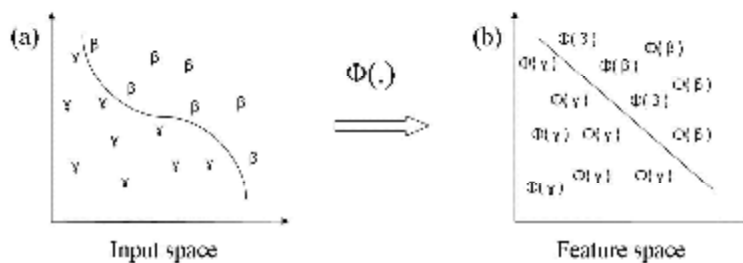
در تعیین تابع $f(x)$ باید همیشه مقدار خروجی به گونه‌ای تعیین شوند که به طور هم‌زمان، میزان کمترین انحراف ε تعیین شده و تابع، مقدار مناسب w را اختیار کند. با حل رابطه بهینه‌سازی زیر دست‌یابی به چنین شرایطی محقق می‌شود.

$$\text{Minimize } \frac{1}{2} w^2 \quad \text{subject to } \begin{cases} y_i - (w, x_i) - b \leq \varepsilon \\ (w, x_i) + b - y_i \leq \varepsilon \end{cases} \quad (2)$$

گاهی ممکن است نتوان میزان انحراف ε را در نظر گرفت و باید مقدار انحراف از همان مقدار ε را نیز قابل قبول فرض کرد. این میزان انحراف اضافه با در نظر گرفتن متغیرهای (ξ_i, ξ_i') در رابطه بهینه‌سازی (۲) اعمال می‌شود که سبب رابطه بهینه‌سازی ذیل می‌شود (وینیک، ۱۹۹۵).

$$\text{Minimize } \frac{1}{2} w^2 + C \sum_{i=1}^k (x_i + x_i') \quad (3)$$

$$\text{subject to } \begin{cases} y_i - (w, x_i) - b \leq \varepsilon + x_i \\ (w, x_i) + b - y_i \leq \varepsilon + x_i \\ x_i, x_i' \geq 0 - \text{for all } i = 1, 2, \dots, K \end{cases}$$



شکل ۱- مفهوم تبدیل از فضای ورودی به فضای مشخصه (دیباک و همکاران، ۲۰۰۱)

تابع موجک، تابعی است که دو ویژگی مهم نوسانی بودن و کوتاه‌مدت بودن را دارد. $\psi(x)$ ، تابعی موجک است اگر و فقط اگر تبدیل فوریه آن $\Psi(\omega)$ ، شرط زیر را ارضا کند (مالات، ۱۹۹۸).

$$\int_{-\infty}^{+\infty} \frac{|\Psi(\omega)|}{|\omega|^2} d\omega < +\infty \quad (5)$$

تبدیل موجک

تابع تبدیل موجک (شکل ۲) قابلیت تجزیه سری زمانی به چندین زیرسری زمانی با مقیاس‌های مختلف را دارد و با بررسی زیرسری‌های زمانی منتجه از سری زمانی کلی، رفتار کوچک مقیاس و بزرگ مقیاس یک فرآیند هیدرولوژیکی را آنالیز می‌کند.

$CWT(a,b) =$

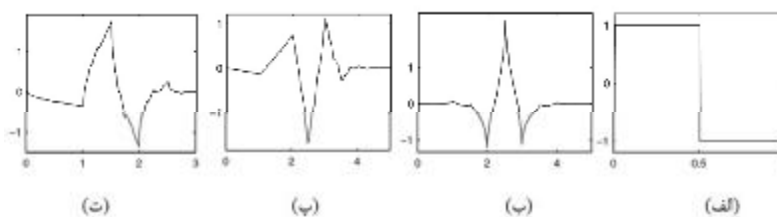
$$Wf(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} f(x) \psi\left(\frac{x-b}{a}\right) dx = \int_{-\infty}^{+\infty} f(x) \psi_{a,b}(x) dx \quad (7)$$
 توابع موجک دارای انواع بسیاری هستند که مهم‌ترین و پرکاربردترین آن‌ها شامل تابع موجک Haar، Meyer، Mexican hat، Morle، Symlet، Daubechies می‌باشند که در شکل ۳ نمودار تابع چهار نمونه از موجک‌ها نشان داده شده است (مالات، ۱۹۹۸).

این شرط با عنوان شرط پذیرفتگی برای موجک $\psi(x)$ شناخته می‌شود. $\psi(x)$ تابع موجک مادر است (رابطه ۶) که ضرایب a و b در آن نقش انتقال و مقیاس سیگنال مورد نظر را دارند (مالات، ۱۹۹۸).

$$\psi_{a,b}(x) = \frac{1}{\sqrt{a}} \psi\left(\frac{x-b}{a}\right) \quad (6)$$
 در نهایت ضریب‌های موجک در هر نقطه از سیگنال (b) و برای هر مقدار از مقیاس (a) با رابطه (۷) قابل محاسبه است (مالات، ۱۹۹۸):



شکل ۲- طرح شماتیک تبدیل موجک



شکل ۳- نمودار (الف) تابع موجک Haar (ب) تابع موجک Coif1 (پ) تابع موجک Sym3 (ت) تابع موجک Db2

در این مدل‌سازی برای سنجش میزان صحت و دقت از روابط آماری ضریب تبیین R^2 و جذر میانگین مربع خطا $(RMSE)^2$ در مرحله‌ی واسنجی و صحت‌سنجی به شرح روابط (۸) و (۹) استفاده می‌شود. بدین منظور داده‌ها به دو گروه تقسیم می‌شوند: ۷۵٪ کل داده‌ها برای واسنجی^۳ و ۲۵٪ باقیمانده برای صحت‌سنجی^۴ مدل ترکیبی استفاده می‌شوند.

$$R^2 = 1 - \frac{\sum_{i=1}^N (Q_{obs_i} - Q_{com_i})^2}{\sum_{i=1}^N (Q_{obs_i} - \bar{Q}_{obs})^2} \quad (8)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (Q_{obs_i} - Q_{com_i})^2}{N}} \quad (9)$$

به طوری که R^2 ، $RMSE$ ، N ، Q_{obs_i} ، Q_{com_i} و \bar{Q}_{obs} به ترتیب ضریب تبیین، جذر میانگین مربع خطا، تعداد

مدل ترکیبی ماشین بردار پشتیبان موجکی (WSVM)

معماری مدل ترکیبی WSVM از دو بخش تبدیل موجک و SVM تشکیل شده است که تبدیل موجک نقش پیش‌پردازش داده‌های خام و مدل SVM نقش مدل‌سازی را دارد (شکل ۴).

در روش پیشنهاد شده ابتدا سری‌های بارش $(I_{(t)})$ و رواناب $(Q_{(t)})$ به زیرسری‌هایی با درجه تجزیه مختلف تجزیه می‌شوند. اگر i و j به ترتیب درجه‌ی تجزیه سری‌های بارش و رواناب در نظر گرفته شوند، تعداد داده‌های ورودی برابر با $i+j+2$ خواهد بود چرا که مدل از دو سری زمانی استفاده می‌کند. بنابراین هر سری زمانی به $i+1$ و $j+1$ زیرسری تجزیه می‌شود. در مرحله بعد داده‌های تجزیه شده برای مدل‌سازی به عنوان ورودی مدل SVM به کار گرفته می‌شوند. در این پژوهش برای انجام مدل‌سازی از زبان برنامه‌نویسی MATLAB (R2011b) استفاده شده است.

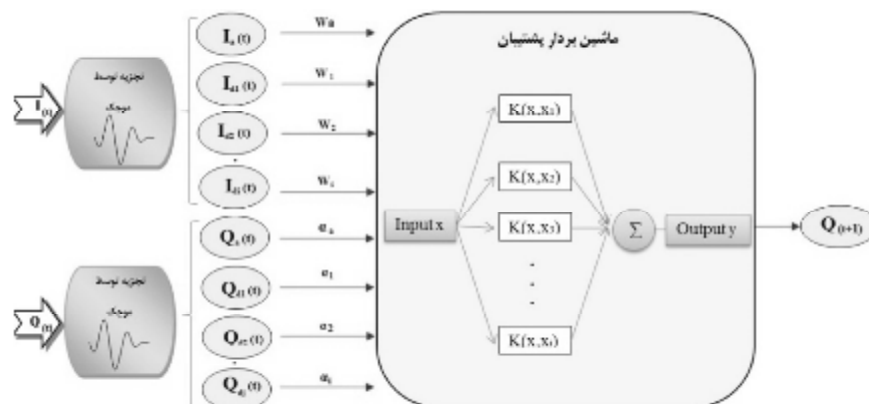
1- Determination Coefficient

2- Root Mean Square Error

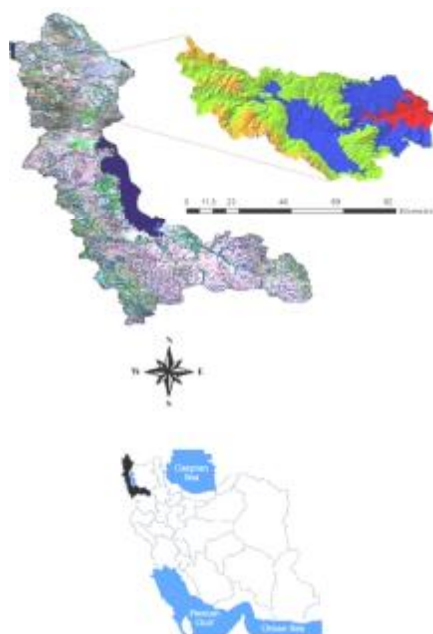
3- Calibration

4- Verification

داده‌های مشاهداتی، داده‌های مشاهداتی، داده‌های پیش‌بینی شده (محاسباتی) و میانگین داده‌های مشاهداتی است.



شکل ۴- ساختار مدل ترکیبی WSVM



شکل ۵- حوضه آقچای (موقعیت جغرافیایی و نقشه DEM)

حوضه مطالعاتی و داده‌ها

حوضه آقچای

حوضه آقچای در شمال شهرستان خوی و جنوب شهرستان سیه چشمه در استان آذربایجان غربی قرار دارد. این حوضه بین $38^{\circ}40'$ و $39^{\circ}30'$ شمال عرض جغرافیایی و $44^{\circ}10'$ و $44^{\circ}57'$ شرق طول جغرافیایی است. مساحت این حوضه 1440 کیلومتر مربع با طول آبراهه به اندازه $64/88$ کیلومتر می‌باشد (شکل ۵). رقوم ارتفاعی این حوضه بین مقادیر 1168 متر و 3280 متر بالای سطح دریا متغیر است. شیب توپوگرافی این منطقه 25% است. مهم‌ترین رودخانه‌های این حوضه قره سو و آقچای هستند. داده‌های بارش و رواناب روزانه 12 سال (از سال 1995 تا سال 2007) مربوط به این حوضه برای فرآیند مدل‌سازی استفاده شده است (8 سال اول برای واسنجی و 4 سال بعدی برای صحت‌سنجی). ویژگی‌های آماری بارش و رواناب روزانه مربوط به این حوضه در جدول ۱ ارائه شده است.

جدول ۱- مشخصه‌های آماری داده‌های بارش و رواناب روزانه برای حوضه مورد مطالعه

سری‌های زمانی رواناب (m^3/s)				
واریانس	میانگین	کمینه	بیشینه	
۳۳/۸۱	۴/۰۳	۰/۱۲	۸۵/۴۰	واسنجی
۳۳/۵۴	۴/۷۵	۰/۱۰	۶۳/۳۰	صحت‌سنجی
سری‌های زمانی بارش (mm)				
واریانس	میانگین	کمینه	بیشینه	
۶/۷۵	۰/۶۹	۰	۲۹	واسنجی
۸/۹۶	۰/۹۱	۰	۳۰	صحت‌سنجی

نتایج مدل‌سازی و بحث

مدل SVM از پارامترهای ساختاری مختلفی همچون تابع کرنل تشکیل شده است که انواع مختلفی دارد. بنابراین در گام نخست، مدل SVM منفرد در نظر گرفته شده و کارآمدترین تابع کرنل انتخاب می‌شود تا در مدل ترکیبی به کار گرفته شود. سپس این مدل با آنالیز موجک ترکیب شده و نتایج حاصل از آن همچون انتخاب بهترین تابع موجک و بهترین درجه تجزیه، صحت‌سنجی می‌شود.

مدل منفرد ماشین بردار پشتیبان (SVM)

یک نکته مهم و اساسی در ساختار الگوریتم SVM تابع کرنل یا هسته است. تابع هسته انواع مختلفی دارد. در این پژوهش سه نوع تابع هسته ارزیابی شده است که شامل RBF-kernel، Poly-kernel و MLP-kernel است. در گام نخست هدف انتخاب بهترین تابع هسته از نظر کارایی در مدل‌سازی در مقایسه با سایر توابع هسته است. بدین منظور همان‌گونه که در جدول ۲ نشان داده شده، مدل‌سازی SVM با سه نوع تابع هسته روی داده‌های بارش و رواناب در زمان $(Q(t), I(t))$ برای حوضه آقچای انجام می‌شود. همچنین R^2 و RMSE به عنوان معیاری برای تعیین کارایی مدل در واسنجی و صحت‌سنجی استفاده می‌شوند.

مطابق جدول ۲، براساس مقدار R^2 و RMSE در صحت‌سنجی (شبه‌سازی)، تابع هسته RBF-kernel در مقایسه با Poly و MLP-kernel کارایی بالاتری دارد. در مورد تابع هسته RBF-kernel باید اشاره کرد که در یادگیری به زبان ماشین، تابع پایه‌ای شعاعی (گوسی) کرنل یا هسته یا همان RBF-kernel یک تابع هسته‌ی مشهور است که در طبقه‌بندی به روش SVM استفاده می‌شود (چانگ و همکاران، ۲۰۱۰). روش کار الگوریتم SVM به این صورت است که داده‌های بارش و رواناب به عنوان

داده‌های ورودی در زمان‌های مختلف با الگوهای ترکیبی متفاوت وارد الگوریتم SVM شده و به این صورت پس از وقوع عملیات مدل‌سازی، رواناب در زمان $t+1$ ($Q(t+1)$) شبیه‌سازی یا پیش‌بینی می‌شود. به بیانی دیگر الگوریتم SVM در مدل‌سازی فرآیندهای بارش و رواناب روزانه عملکردی خود همبسته دارد به طوری که برای پیش‌بینی رواناب آینده ($Q(t+1)$) از داده‌های زمان t و $t-1$ ($I(t), I(t-1), I(t-2), \dots, Q(t), Q(t-1), Q(t-2), \dots$) استفاده می‌شود. در این پژوهش از الگوریتم مارکوف برای تعیین وابستگی زمانی داده‌های بارش و رواناب استفاده شده است. بدین ترتیب که برای پیش‌بینی رواناب $Q(t+1)$ در زمان $t+1$ از ترکیب ورودی‌های بارش و رواناب I و Q به شرح ترکیب‌های زیر استفاده شده است.

Comb. (1): $Q(t), I(t)$

Comb. (2): $Q(t), Q(t-1), I(t), I(t-1)$

Comb. (3): $Q(t), Q(t-1), Q(t-2), I(t), I(t-1), I(t-2)$

Comb. (4): $Q(t), Q(t-1), Q(t-2), Q(t-3), I(t), I(t-1), I(t-2), I(t-3)$

Comb. (5): $Q(t), Q(t-1), Q(t-2), Q(t-3), Q(t-4), I(t), I(t-1), I(t-2), I(t-3), I(t-4)$

در این مرحله هدف انتخاب بهترین ترکیب داده‌های ورودی در مقایسه با سایر ترکیب‌ها از نظر کارایی در مدل‌سازی است. بدین منظور، مدل‌سازی با تابع هسته برگزیده در مرحله قبل روی ترکیب‌های مختلف از داده‌های ورودی انجام می‌شود. سپس با توجه به مقادیر R^2 و RMSE دقت مدل‌سازی در مراحل واسنجی و صحت‌سنجی مشخص می‌شود.

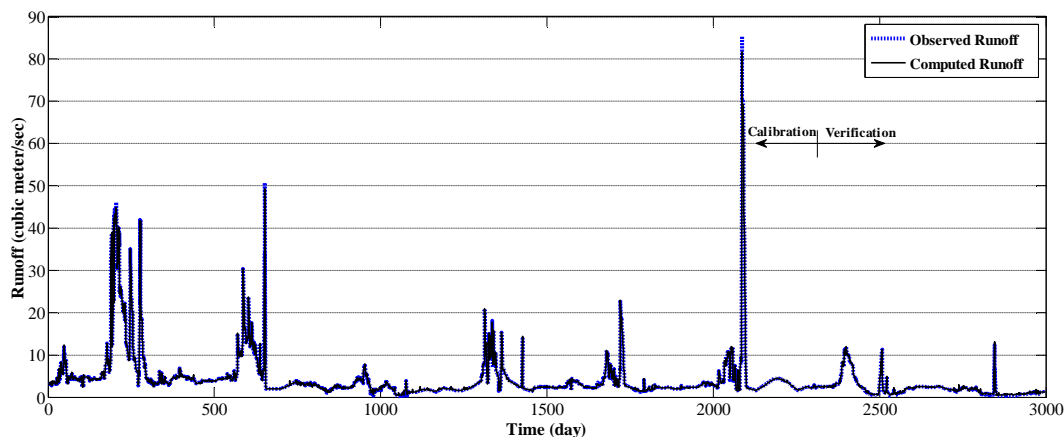
نتایج حاصله نشان از آن هستند که دومین ساختار ترکیب داده‌ها (Comb (2)) بهترین کارایی را در مدل‌سازی از خود نشان می‌دهد (جدول ۳). مقایسه‌ی داده‌های مشاهداتی و محاسباتی رواناب که از تابع هسته RBF-kernel و Comb(2) به دست آمده، در شکل ۶ برای مراحل واسنجی و صحت‌سنجی نشان داده شده است.

جدول ۲- نتایج حاصل از سری‌های زمانی حوضه آقچای با اطلاعات مربوط به $Q(t)$ و $I(t)$ برای انتخاب بهترین تابع هسته

تابع هسته	R^2		RMSE (m ³ /s)	
	واسنجی	صحت‌سنجی	واسنجی	صحت‌سنجی
Poly-kernel	۰/۵۲	۰/۶۴	۰/۵۶	۰/۳۷
RBF-kernel	۰/۶۴	۰/۶۵	۰/۱۶	۰/۱۹
MLP-kernel	۰/۵۵	۰/۵۶	۰/۴۴	۰/۵۴

جدول ۳- نتایج حاصل از سری های زمانی بارش و رواناب حوضه آقچای با مدل سازی تحت ترکیب ورودی های مختلف با تابع هسته

ترکیب ورودی ها	R ²		RMSE (m ³ /s)	
	واسنجی	صحت سنجی	واسنجی	صحت سنجی
Comb. (1)	۰/۵۳	۰/۵۹	۰/۲۲	۰/۸۶
Comb. (2)	۰/۶۴	۰/۶۹	۰/۱۵	۰/۷۹
Comb. (3)	۰/۶۲	۰/۴۲	۰/۲۷	۱/۲۲
Comb. (4)	۰/۶۲	۰/۳۳	۰/۲۹	۱/۴۹
Comb. (5)	۰/۶۳	۰/۲۵	۰/۳۱	۱/۷۴



شکل ۶- داده های مشاهداتی و محاسباتی رواناب به دست آمده از تابع هسته RBF-kernel و Comb (2)

داده های محاسبه شده با داده های مشاهداتی انجام می شود. نتایج حاصل از مدل سازی بر اساس بهترین تابع هسته (RBF-kernel) برای انتخاب بهترین تابع موجک و درجه ی تجزیه در جدول ۴ آورده شده است. در این جدول i و j به ترتیب درجه ی تجزیه سری های بارش و رواناب هستند که به صورت یکسان در نظر گرفته می شوند. با مقایسه نتایج جدول ۴ با جدول های ۲ و ۳ می توان چنین نتیجه گیری نمود که مدل ترکیبی WSVM به دلیل عمل پیش پردازش روی داده های ورودی نتایج دقیق تری را نسبت به مدل ساده SVM ارائه کرده است. به بیان کامل تر وقتی تابع موجک داده های بارش و رواناب را به داده های با مقیاس زمانی مختلف تفکیک می کند، عمل طبقه بندی و رگرسیون داده ها با الگوریتم SVM با خطای کمتری انجام می شود.

با توجه به بالاترین مقدار ضریب تبیین در مراحل واسنجی و صحت سنجی که به ترتیب برابر ۰/۹۶ و ۰/۹۲ است، بهترین نتیجه برای ساختار موجک مادر Db2 با درجه ی تجزیه سه است. سری زمانی مربوط به داده های مشاهداتی و محاسباتی در مدل سازی به روش ترکیبی WSVM برای

مدل ترکیبی ماشین بردار پشتیبان موجکی (WSVM)

در این مرحله برای سنجش اثر آنالیز موجک بر فرآیند مدل سازی، داده های پیش پردازش شده به عنوان ورودی های مدل SVM استفاده می شوند. بدین منظور از تبدیل موجک به صورت مجزا استفاده می شود. تبدیل موجک داده ها را در مقیاس های زمانی مختلف (درجه تجزیه) ارزیابی می کند. بنابراین سری های زمانی کوتاه و بلند مقیاس در این فرآیند از هم جدا می شوند. در این پژوهش تلاش شده تا اثر استفاده از توابع موجک مختلف با درجه تجزیه مختلف روی کارایی مدل مشخص شود. بنابراین سری های زمانی با درجه تجزیه ۱ تا ۳ با سه نوع تابع موجک تجزیه می شوند. تابع موجک Haar، تابع موجک Db2 و تابع موجک Sym3 (مالات، ۱۹۹۸) توابع موجکی هستند که در این پژوهش از آن ها استفاده شده است. در مدل ترکیبی روش کار بدین شرح است که سری های تجزیه شده با تابع موجک وارد مدل SVM شده و پس از واسنجی مقدار رواناب را در یک روز جلوتر ($Q(t+1)$) به عنوان خروجی پیش بینی می کند و سپس صحت سنجی

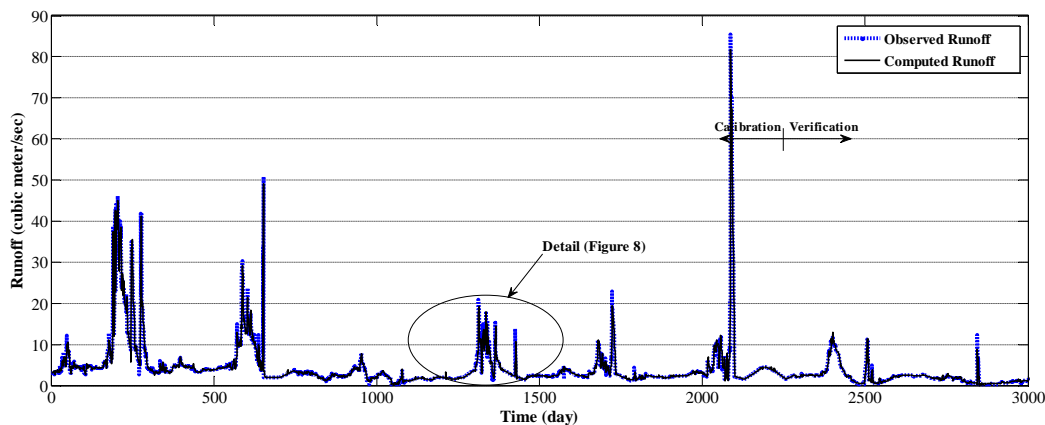
دقت بالای WSVM در مدل‌سازی داده‌های حوضه الیور است. همچنین برای مقایسه‌ی مدل ترکیبی WSVM با سایر مدل‌های ترکیبی مانند شبکه عصبی مصنوعی موجکی (WANN)، داده‌های بارش و رواناب حوضه آقچای نیز تحت مدل‌سازی WANN قرار گرفته و بهترین نتیجه مدل‌سازی که مانند مدل WSVM با تابع موجک Db2 و درجه تجزیه سه به دست می‌آید، در جدول ۶ نشان داده شده است. نتایج حاصله گویای این واقعیت است که مدل ترکیبی WSVM با ضریب تبیین ۰/۹۲ در مقایسه با مدل ترکیبی WANN با ضریب تبیین ۰/۸۷ دارای کارایی و دقت بالاتری در مدل‌سازی فرآیند بارش - رواناب است. در نتیجه مدل ترکیبی WSVM مدلی قدرتمند برای مدل‌سازی فرآیندهای هیدرولوژیکی مانند بارش و رواناب معرفی می‌شود که این خود سبب پیشرفت زیادی در پیش‌بینی سیلاب و همچنین مدیریت منابع آبی در کشور می‌شود.

بهترین تابع موجک (Db2) و درجه‌ی تجزیه سه در شکل ۷ نشان داده شده است. برای درک بهتر از داده‌های محاسباتی و مشاهداتی، ناحیه‌ی مشخص شده در شکل ۷، به صورت بزرگ‌نمایی در شکل ۸ ارائه شده است. نکته‌ی مهمی که از شکل‌های ۷ و ۸ استنباط می‌شود توانایی بالای مدل ترکیبی WSVM در محاسبه نقاط بیشینه سری زمانی در قیاس با مدل SVM است. همچنین یک نکته‌ی مهم در مدل‌سازی WSVM عدم نیاز به کنترل ترکیبات مختلف (Comb) به شیوه‌ی مارکوف است زیرا مدل SVM با استفاده از آنالیز موجک می‌تواند الگوها و روندهای حاکم بر پدیده را به خوبی رصد کند.

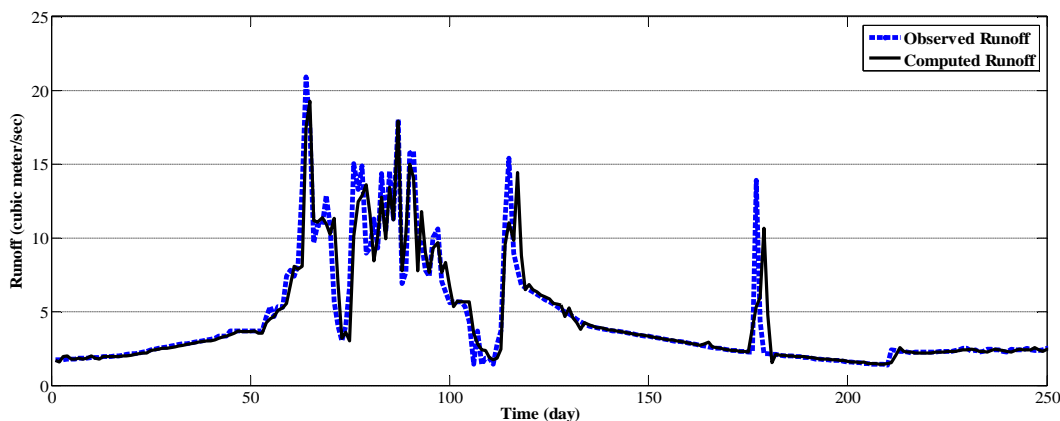
برای اثبات سازگاری مدل WSVM با داده‌های سایر حوضه‌ها، داده‌های بارش و رواناب حوضه الیور در ایالت کالیفرنیا استفاده می‌شود. نتایج این مدل‌سازی در جدول ۵ آورده شده است. همان‌طور که پیداست بیشترین ضریب تبیین در صحت‌سنجی این حوضه ۰/۹۱ و این نشان‌دهنده

جدول ۴- نتایج حاصل از مدل ترکیبی WSVM با تابع هسته RBF-kernel با توابع مختلف موجک و درجه‌های تجزیه متفاوت

نوع تابع موجک	درجه تجزیه (i=j)	صحت‌سنجی		واسنجی	
		R^2	صحت‌سنجی	$RMSE (m^3/s)$	واسنجی
Haar	۱	۰/۹۲	۰/۹۰	۰/۱۶	۰/۶۱
Haar	۲	۰/۹۳	۰/۸۲	۰/۱۵	۰/۸۵
Haar	۳	۰/۹۳	۰/۷۹	۰/۱۲	۰/۹۳
Db2	۱	۰/۹۲	۰/۷۷	۰/۱۳	۰/۹۹
Db2	۲	۰/۹۳	۰/۸۳	۰/۱۵	۰/۸۳
Db2	۳	۰/۹۶	۰/۹۲	۰/۰۹	۰/۳۵
Sym3	۱	۰/۹۲	۰/۹۱	۰/۱۹	۰/۶۱
Sym3	۲	۰/۹۴	۰/۸۲	۰/۱۵	۰/۸۵
Sym3	۳	۰/۹۴	۰/۷۹	۰/۱۴	۰/۹۳



شکل ۷- مقایسه داده‌های محاسباتی و مشاهداتی مربوط به تابع موجک Db2 و درجه تجزیه ۳



شکل ۸- مقایسه داده‌های محاسباتی و مشاهداتی در ناحیه مشخص شکل ۶

جدول ۵- نتایج حاصل از مدل ترکیبی WSVM با تابع هسته RBF-kernel با تابع موجک Db2 و درجه‌های تجزیه متفاوت مربوط به حوضه الریور

نوع تابع موجک	درجه تجزیه (i=j)	واسنجی		واسنجی	
		R ²	صحت‌سنجی	RMSE(m ³ /s)	صحت‌سنجی
Db2	۱	۰/۹۰	۰/۷۶	۶/۹۵	۲۲۵/۵۱
Db2	۲	۰/۹۰	۰/۸۰	۸/۸۲	۱۹۹/۶۳
Db2	۳	۰/۹۳	۰/۹۱	۶/۹۵	۱۸۸/۶۴
Db2	۴	۰/۹۲	۰/۸۱	۵/۶۵	۲۲۹/۳۷

جدول ۶- مقایسه نتایج مدل‌سازی مدل WSVM با مدل WANN

نام مدل	نوع مدل	R ² (واسنجی)	R ² (صحت‌سنجی)
WSVM	ترکیبی	۰/۹۶	۰/۹۲
WANN	ترکیبی	۰/۹۳	۰/۸۷

توجه: در این جدول بهترین نتایج از هر مدل نمایش داده شده است.

نتیجه‌گیری

در این پژوهش مدل ترکیبی ماشین بردار پشتیبان موجکی (WSVM) برای مدل‌سازی سری زمانی بارش رواناب ارائه شد. بدین منظور سری‌های زمانی بارش و رواناب مربوط به حوضه‌ی آقچای با تبدیل موجک به زیرسری‌هایی تجزیه شد و سپس این زیرسری‌های تجزیه شده به عنوان داده‌های ورودی در مدل SVM برای پیش‌بینی رواناب در یک گام زمانی جلوتر استفاده شد. نتایج نشان داد که تابع هسته RBF-Kernel نسبت به سایر توابع دارای دقت و کارایی بالاتری است. نکته مهم این پژوهش، بهره‌گیری از توابع تبدیل موجک برای رفع مشکل استفاده از ترکیب زیاد و ویژگی خود همبستگی مدل SVM است. بدین منظور از سه نوع تابع موجک (Sym3 و Db2، Haar) با

مدل موردنظر و درجات آزادی یک تا سه برای مدل‌سازی استفاده شد. نتایج مدل‌سازی نشان داد که تابع موجک Db2 و درجه تجزیه سه به عنوان مناسب‌ترین گزینه‌ها از نظر کارایی در مدل ترکیبی هستند. همچنین نشان داده شد که یکی از توانایی‌های ویژه مدل ترکیبی WSVM نسبت به مدل منفرد، محاسبه‌ی نقاط بیشینه سری زمانی رواناب است. در نهایت نتایج مطلوب مدل‌سازی WSVM برای داده‌های حوضه الریور و همچنین دقت بالای این مدل در مقایسه با مدل ترکیبی WANN به اثبات رسید. به دلیل قابلیت ویژه مدل ترکیبی پیشنهاد می‌شود که از این مدل برای مدل‌سازی فرآیندهای هیدرولوژیکی مانند رسوب‌گذاری و خشکسالی نیز استفاده شود.

13. Vapnik V. N. and Cortes C. 1995. Support vector networks. *Machine Learning*. 20: 273-297.
14. Vapnik V. N. 1995. *The nature of statistical learning theory*. Springer, New York.
15. Wang L. 2005. *Support vector machines: theory and applications*. Springer-Berlin 177.
16. Zhang B. L. and Dong Z. Y. 2001. An adaptive neural wavelet model for short term load forecasting. *Electric Power Systems Research*. 59: 121-129.

منابع

1. Abe S. 2005. *Support vector machines for pattern classification* (Vol. 53). London: Springer.
2. Burges C. J. C. 1999. *Geometry and invariance in kernel based methods in advance in kernel methods-support vector learning*. Cambridge, MA: MIT Press. 89-116.
3. Chang Y. W. Hsieh C. J. Chang K. W. Ringgaard M. and Lin C. J. 2010. Training and testing low-degree polynomial data mappings via linear support vector machine. *Journal of Machine Learning Research*. 11: 1471-1490.
4. Choy K. Y. and Chan C. W. 2003. Modelling of river discharges and rainfall using radial basis function networks based on support vector regression. *International Journal of Systems Science*. 34(14-15): 763-773.
5. Dibike Y. B. Velickov S. Solomatine D. and Abbott M. B. 2001. Model induction with support vector machines: Introduction and applications. *Journal of Computing in Civil Engineering*. 15 (3): 208-216.
6. Garbrecht J. D. 2006. Comparison of three alternative artificial neural network designs for monthly rainfall-runoff simulation. *Journal of Hydrology*. 11(5): 502-505.
7. Lin G. Chen G. Huang P. and Chou Y. 2009. Support vector machine-based models for hourly reservoir inflow forecasting during typhoon- warning periods. *Journal of Hydrology*. 372: 17-29.
8. Mallat S. G. 1998. *A wavelet tour of signal processing*. Second ed. Academic Press. San Diego.
9. Nourani V. Komasi M. Taghi Alami M. 2012. Hybrid wavelet-genetic programming approach to optimize artificial neural network modeling of rainfall-runoff process. *Journal of Hydrologic Engineering* © ASCE. 17(6): 724-741.
10. Rajaei T. Mirbagheri S. A. Nourani V. and Alikhani A. 2010. Prediction of daily suspended sediment load using wavelet and neuro-fuzzy combined model. *Journal of Environmental Science and Technology*. 7(1): 93-110.
11. Rajurkar M. P. Kothiyari U. C. and Chaube U. C. 2004. Modeling of the daily rainfall-runoff relationship with artificial neural network. *Journal of Hydrology*. 285(1-4): 96-113.
12. Salas J. D. Delleur J. W. Yevjevich V. and Lane W. L. 1980. *Applied modeling of hydrological time series*. Water resources publications, Denver.