

مقایسه روش های برنامه ریزی ژنتیک و ماشین بردار پشتیبان در پیش بینی جریان روزانه رودخانه (مطالعه موردی: رودخانه باراندوزچای)

فرشاد احمدی^{۱*} - فریدون رادمش^۲ - رسول میرعباسی نجف آبادی^۳

تاریخ دریافت: ۱۳۹۲/۱۱/۲۸

تاریخ پذیرش: ۱۳۹۳/۶/۱۰

چکیده

پیش بینی دقیق جریان رودخانه‌ها در مدیریت منابع آب از اهمیت بسزایی برخوردار است. در مطالعه حاضر به منظور پیش‌بینی جریان رودخانه باراندوزچای از دو روش برنامه‌ریزی ژنتیک (GP) و ماشین بردار پشتیبان (SVM) استفاده شد. داده‌های جریان روزانه این رودخانه در ایستگاه دیزج در خلال سال‌های ۱۳۸۵ تا ۱۳۸۹ برای ایجاد مدل استفاده شد که ۸۰ درصد داده‌ها برای آموزش و ۲۰ درصد برای تست مدل بکار رفت. نتایج نشان داد که در هر دو روش، مدل‌های شامل جریان یک، دو و سه روز قبل بالاترین دقت را در مرحله صحت‌سنجی داشتند. همچنین دقت هر دو مدل با افزایش مقادیر دبی کاهش می‌یابد. مقایسه نتایج دو مدل نشان داد که گرچه دقت روش برنامه‌ریزی ژنتیک با $R=0.978$ و $RMSE=1/66 (m^3/s)$ نسبت به روش ماشین بردار پشتیبان با $R=0.976$ و $RMSE=1/80 (m^3/s)$ اندکی بیشتر بود، اما روش SVM به مراتب ساده‌تر از روش GP می‌باشد و این روش می‌تواند به عنوان یک روش کاربردی برای پیش‌بینی جریان روزانه بکار رود.

واژه های کلیدی: برنامه‌ریزی ژنتیک، پیش‌بینی جریان روزانه، ماشین بردار پشتیبان، رودخانه باراندوزچای

مقدمه

زمانی مورد بررسی و به ویژه فواصل زمانی کوتاه‌تر از زمان تمرکز حوضه را داراست. مقایسه نتایج با مطالعه دیگری که در آن روش‌های خودهمبستگی و فیلترکالمن استفاده شده بود، مشخص نمود که برنامه ریزی ژنتیک ابزار مناسب‌تری برای پیش‌بینی رواناب است. همچنین بر طبق تحقیقات آیتک و همکاران (۸) روش برنامه‌ریزی ژنتیک یک روش مناسب و علمی در پیش‌بینی روابط بارش - رواناب می‌باشد. گوون (۱۲) برای پیش‌بینی دبی جریان روزانه رودخانه شویل کیل^۵ در ایالات متحده از روش برنامه‌ریزی ژنتیک خطی استفاده و نتایج حاصل را با دو الگو از روش‌های شبکه عصبی مورد مقایسه قرارداد. وی نشان داد که هر دو روش نتایج قابل قبولی داشته اند، اما روش برنامه‌ریزی ژنتیک خطی از دقت بالاتری نسبت به روش‌های شبکه عصبی برخوردار است. قربانی و همکاران (۱۱) عملکرد سه روش برنامه ریزی ژنتیک، شبکه عصبی مصنوعی^۶ و نروفازی^۷ را در روندیابی سیلاب رودخانه قزل ایرماق ترکیه مورد ارزیابی قرار دادند. نتایج مطالعه آنها نشان داد که از بین سه روش مذکور مدل برنامه ریزی ژنتیک با دقت بیشتری هیدروگراف خروجی

نیاز روزافزون به آب سبب گردیده است که برنامه‌ریزی‌های مدیریتی به منظور کنترل مصرف آب در آینده از اهمیت بیشتری برخوردار باشد. با پیش‌بینی نمودن جریان رودخانه‌ها علاوه بر مدیریت بهره‌برداری از منابع آب، می‌توان حوادث طبیعی نظیر سیل و خشکسالی را نیز پیش‌بینی و مهار نمود. برنامه‌ریزی ژنتیک^۴ شاخه‌ای از الگوریتم‌های تکاملی است که توانایی مدل‌سازی فرایندهای کاملاً غیرخطی و پویا را دارند. تاکنون محققین مختلفی در سراسر جهان با روش‌های مختلف اقدام به پیش‌بینی جریان رودخانه‌ها در مقیاس‌های زمانی ساعتی، روزانه و ماهانه نموده‌اند. خو و همکاران (۱۴) رواناب ساعتی حوضه آرگوال فرانسه را با استفاده از روش برنامه‌ریزی ژنتیک، پیش‌بینی نمودند. بررسی نتایج نشان داد که برنامه‌ریزی ژنتیک توانایی پیش‌بینی دقیق رواناب را در تمام فواصل

۱ و ۲- دانشجوی دکتری و دانشیار گروه مهندسی منابع آب، دانشکده مهندسی

علوم آب، دانشگاه شهید چمران اهواز

(*- نویسنده مسئول: (Email: Farshad.paper@yahoo.com)

۳- استادیار گروه مهندسی آب، دانشگاه شهرکرد

4 - Genetic Programming

5 - Schuykill

6- Artificial Neural Network (ANN)

7- Neuro-fuzzy

روش رگرسیون ماشین بردار پشتیبان (SVR) و شبکه های موجک^۵ (WN) استفاده کردند. مقایسه نتایج نشان داد که گرچه هر دو روش از دقت خوبی در پیش بینی جریان روزانه برخوردار بودند، اما دقت روش شبکه های موجک اندکی بیشتر بود.

با عنایت به موارد فوق می توان دریافت که تخمین پارامترهای هیدرولوژیکی هم چون پیش بینی جریان رودخانه ها از دیرباز مورد توجه محققین امر بوده و بدین منظور روش های متعددی از جمله مدل های تجربی- نیمه تجربی، سری های زمانی و مدل های هوشمند توسعه یافته اند که در این میان مدل های هوشمند با الهام گرفتن از طبیعت قادر به تخمین پارامترهای مربوط به پدیده های طبیعی با دقت قابل قبول می باشند و نیز دقت قابل توجهی نسبت به سایر روش ها دارند. لذا در این تحقیق سعی گردیده است که عملکرد دو مدل هوشمند برنامه ریزی ژنتیک (GP) و ماشین بردار پشتیبان (SVM) در پیش-بینی جریان روزانه رودخانه باراندوزچای مورد بررسی قرار گرفته و اهداف زیر دنبال شود: ۱) توسعه مدل های مناسب GP و SVM برای پیش بینی جریان روزانه رودخانه باراندوزچای، ۲) استخراج ساختار درختی و رابطه ریاضی حاکم بر دبی روزانه رودخانه مورد مطالعه با استفاده از روش GP، ۳) تعیین مقدار بهینه ضرایب سه گانه مدل SVM و پیش بینی جریان روزانه.

مواد و روش ها

منطقه مورد مطالعه

در این مطالعه، از داده های دبی جریان روزانه ایستگاه هیدرومتری دیزج واقع بر رودخانه باراندوزچای ارومیه در دوره آماری ۱۳۸۵ تا ۱۳۸۹ استفاده گردید. مساحت حوضه باراندوزچای در ایستگاه هیدرومتری دیزج ۶۶۰/۷۱ کیلومتر مربع است. این حوضه در شمال غرب کشور بین دریاچه ارومیه و مرز ایران و کشورهای عراق و ترکیه در موقعیت جغرافیایی ۴۴°۴۵' تا ۴۵°۱۴' طول شرقی و ۳۷°۰۶' تا ۳۷°۲۹' عرض شمالی واقع شده است. طول آبراهه اصلی حوضه ۷۵ کیلومتر بوده و حداکثر و حداقل ارتفاع این حوضه به ترتیب ۳۵۰۰ و ۱۲۵۰ متر از سطح آزاد دریاها می باشد. در جدول ۱ مشخصات آماری ایستگاه هیدرومتری دیزج آورده شده است. شکل های ۱ و ۲ نیز به ترتیب نمودار سری زمانی جریان وزانه رودخانه باراندوزچای و موقعیت ایستگاه هیدرومتری دیزج را نشان می دهد.

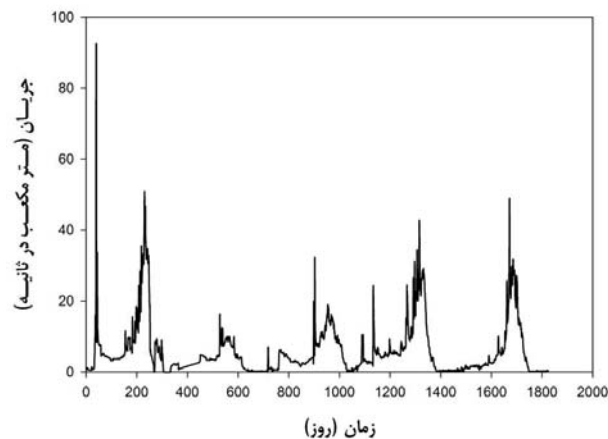
برنامه ریزی ژنتیک

روش برنامه ریزی ژنتیک اولین بار توسط کوزا (۱۵) توسعه داده شد. این روش جزو روش های الگوریتم تکاملی محسوب می شود که مبنای آنها براساس نظریه تکامل داروین استوار است.

را شبیه سازی می کند. ظهیری و عظمت الله (۱۹) برای پیش بینی دبی جریان در مقاطع مرکب از دو روش برنامه ریزی ژنتیک خطی و مدل درختی M5 استفاده کردند. نتایج نشان داد که هرچند هر دو مدل از دقت بالایی برای پیش بینی جریان برخوردار بودند، اما دقت روش برنامه ریزی ژنتیک خطی بالاتر از مدل درختی M5 بود.

ماشین بردار پشتیبان یکی از روش های یادگیری تحت نظارت^۱ است که هم برای دسته بندی و هم رگرسیون قابل استفاده است. این روش توسط وپنیک (۱۷) بر پایه تئوری یادگیری آماری^۲ بنا نهاده شده است. ماشین بردار پشتیبان^۳ روشی برای طبقه بندی دوتائی در فضای ویژگی های دلخواه است و از این روشی مناسب برای مسائل پیش بینی به شمار می رود (۱۶). ماشین بردار پشتیبان در اصل یک دسته بندی کننده دو کلاسه است که کلاس ها را توسط یک مرز خطی از هم جدا می کند. در این روش نزدیک ترین نمونه ها به مرز تصمیم گیری را بردارهای پشتیبان می نامند. این بردارها معادله مرز تصمیم گیری را مشخص می کنند. الگوریتم های شبیه سازی هوشمند کلاسیک مانند شبکه های عصبی مصنوعی، معمولاً قدر مطلق خطا یا مجموع مربعات خطای داده های آموزشی را حداقل می کنند، ولی مدل های SVM، اصل حداقل سازی خطای ساختاری را به کار می گیرند (۱۳). اخیراً این مدل ها در گستره وسیعی از مسائل هیدرولوژیکی و به ویژه پیش بینی جریان رودخانه ها به کار رفته اند. نیک بخت شهبازی (۵) از مدل ماشین بردار پشتیبان سری های زمانی و داده های رودخانه های کشف رود (ایستگاه آق دریند) و هریرود (ایستگاه پل خاتون) در یک دوره ۴۵ ساله (۱۳۳۰ تا ۱۳۷۴) جهت تولید جریان استفاده کردند. نتایج نشان داد که ماشین بردار پشتیبان در حفظ توام خواص ماهیانه و سالیانه جریان بسیار خوب عمل می کند. مقایسه نتایج حاصل با مدل HEC4، نشان داد که مدل بردار پشتیبان روشی بسیار بهتر و دقیق تر در تولید جریان مصنوعی بوده و عملکردی به مراتب بهتر از خود نشان می دهد. یو و همکاران (۱۸) از روش ماشین بردار پشتیبان برای پیش بینی سیلاب ساعتی رودخانه لان یانگ در کشور تایوان استفاده کردند. نتایج نشان داد که این روش برای پیش بینی سیلاب یک ساعت آینده از دقت بالایی برخوردار بود. در مطالعه ای دیگر، یو و همکاران (۱۹) از روش ماشین بردار پشتیبان جهت پیش بینی سیلاب ساعتی رودخانه لان یانگ در تایوان استفاده نمودند. نتایج حاصل حاکی از دقت مناسب و قابل قبول روش مذکور در پیش بینی رویداد های سیلاب در ۱ تا ۶ ساعت بعد داشت. آداموفسکی و پراشر (۶) برای پیش بینی جریان روزانه در حوضه کوهستانی سیانجی^۴ واقع در منطقه هیمالیای هندوستان از دو

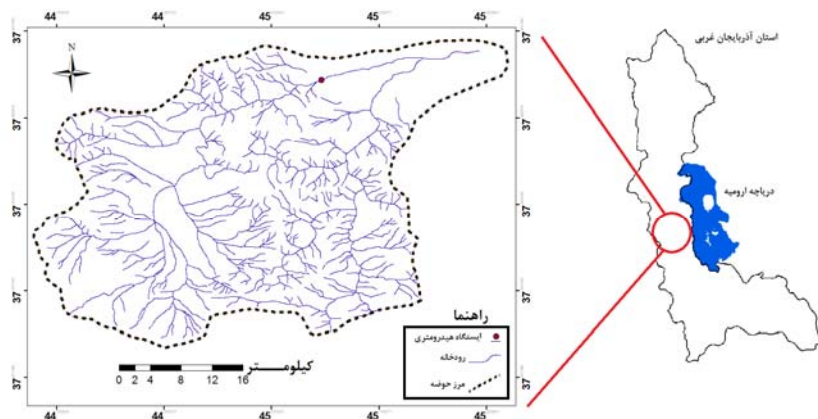
- 1- Supervised learning
- 2- Statistical Learning Theory
- 3- Support Vector Machine (SVM)
- 4- Sianji



شکل ۱- نمودار سری زمانی جریان روزانه رودخانه در دوره آماری ۱۳۸۵-۱۳۸۹

جدول ۱- مشخصات آماری سری جریان روزانه رودخانه باراندوزچای در محل ایستگاه دیزج در دوره آماری ۱۳۸۵-۱۳۸۹

مشخصات آماری دبی روزانه (m ³ /s)	
۵/۹۲	میانگین
۵۷/۹۸	واریانس
۷/۶۱	انحراف معیار
۹۲/۶۰	حداکثر
۰/۰۰	حداقل



شکل ۲- موقعیت حوضه آبریز باراندوزچای و ایستگاه هیدرومتری دیزج

ژنتیک به طبیعت هر یک از افراد بر می‌گردد، به نحوی که افراد در الگوریتم ژنتیک، ردیف‌های خطی با طول ثابت می‌باشند (کروموزم‌ها) ولی در برنامه‌ریزی ژنتیک، همانند شاخه‌های مجزا هستند. همچنین در برنامه‌ریزی ژنتیک بر ساختار درختی مجموعه‌ها تاکید می‌شود ولی الگوریتم ژنتیک، بر اساس سیستم ارقام دودویی عمل می‌نماید. فرآیند گام به گام حل یک مسئله با استفاده از برنامه‌ریزی ژنتیک متشکل از ۵ مرحله به شرح زیر می‌باشد (۴):

الگوریتم‌های یاد شده اقدام به تعریف یک تابع هدف در قالب معیارهای کیفی نموده و سپس تابع یاد شده را برای مقایسه جواب‌های مختلف حل مسئله در یک فرآیند گام به گام تصحیح ساختار داده‌ها به کار می‌گیرند و در نهایت، جواب مناسب را ارائه می‌نمایند. روش برنامه‌ریزی ژنتیک جدیدترین شیوه از بین روش‌های الگوریتم تکاملی می‌باشد که به دلیل دارا بودن دقت کافی، از کاربرد بیشتری برخوردار است (۷). تفاوت اساسی موجود بین الگوریتم ژنتیک و برنامه‌ریزی

$SVM - \nu$ نیز نامیده می شوند و ب) مدل های رگرسیونی SVM نوع دوم که با نام $SVM - \epsilon$ شناخته شده هستند. در این مطالعه $SVM - \epsilon$ به دلیل کاربرد گسترده آن در مسائل رگرسیونی مورد استفاده قرار گرفت. برای این مدل، تابع خطا به صورت زیر تعریف می شود:

$$\frac{1}{2}W^T W + C \sum_{i=1}^N \xi_i + C \sum_{i=1}^N \xi_i^* \quad (2)$$

تابع خطای فوق لازم است که با توجه به محدودیت های زیر کمینه گردد (۱۳):

$$\begin{aligned} W^T \phi(x_i) + b - y_i &\leq \epsilon + \xi_i^* \\ y_i - W^T \phi(x_i) - b &\leq \epsilon + \xi_i \\ \xi_i, \xi_i^* &\geq 0 \end{aligned} \quad (3)$$

که در این روابط C ثابت گنجایش، W بردار ضرایب، W^T ترانهاده بردار ضرایب، ξ_i^* ، ξ_i ضرایب کمبود، b ضریب ثابت، N الگوهای آموزش مدل و ϕ تابع کرنل است. اطلاعات کمی در مورد انتخاب تابع غیرخطی مناسب ϕ در دسترس می باشد. ماشین های بردار پشتیبان برای حل مسائل غیرخطی، ابعاد مساله را از طریق توابع کرنل تغییر می دهند. انتخاب کرنل برای SVM به حجم داده های آموزشی و ابعاد بردار ویژگی بستگی دارد. به عبارت دیگر، باید با توجه به این پارامترها تابع کرنل را انتخاب نمود که توانایی آموزش برای ورودی های مساله را داشته باشد. در عمل چهار نوع کرنل خطی^۳، کرنل چند جمله ای^۴، کرنل تانژانت هیپربولیک^۵ و کرنل گوسی^۶ (RBF) به کار گرفته می شوند. در جدول ۲ معادلات برخی از کرنل های رایج ارائه شده اند.

جدول ۲- توابع کرنل رایج در ماشین های بردار پشتیبان (۱۳)

نوع تابع	تابع کرنل
خطی	$K(x_i, x_j) = x_i^T \cdot x_j$
چند جمله ای	$K(x_i, x_j) = (\gamma x_i^T \cdot x_j + C)^d$
تانژانت هیپربولیک	$K(x_i, x_j) = \tanh(\gamma x_i^T \cdot x_j + C)$
RBF	$K(x_i, x_j) = \exp\left(-\gamma x_i - x_j ^2\right)$

در نهایت، تابع تصمیم رگرسیون بردار پشتیبان غیرخطی به صورت معادله زیر خواهد بود که کنترل کننده میزان نوسان تابع گوسی و همچنین کنترل کننده نتایج پیش بینی و تعمیم دهنده مدل

(۱) انتخاب مجموعه ترمینال: که همان متغیرهای مستقل مسئله و متغیرهای حالت سامانه می باشند. (۲) انتخاب مجموعه توابع: که شامل عملگرهای حسابی، توابع آزمون و توابع بولی^۱ می باشد. (۳) شاخص اندازه گیری دقت مدل که بر مبنای آن می توان مشخص نمود که توانایی مدل در حل یک مسئله خاص تا چه اندازه می باشد. (۴) مولفه های کنترل: مقادیر مولفه های عددی و متغیرهای کیفی که برای کنترل اجرای برنامه ها استفاده می شوند. (۵) شرط توقف اجرای برنامه: که معیاری برای حصول نتیجه و توقف اجرای برنامه می باشد. در مطالعه حاضر از برنامه GeneXpro Tools (۱۰) برای توسعه و اجرای مدل های مبتنی بر برنامه ریزی ژنتیک استفاده شد. فرآیند مدل سازی جریان رودخانه باراندوزچای در تحقیق حاضر به شکل زیر صورت گرفت.

گام اول، انتخاب تابع برازش مناسب می باشد که در این مطالعه تابع جذر میانگین مربعات خطا به عنوان تابع برازش انتخاب گردید. گام دوم، انتخاب مجموعه متغیرهای ورودی و مجموعه توابع به منظور تولید کروموزوم ها می باشد. در مسئله حاضر مجموعه ترمینال ها متشکل از مقادیر جریان رودخانه با تاخیرهای زمانی می باشد. در این مطالعه، از چهار عملگر اصلی که شامل $\{+, -, \times, \div\}$ و نیز توابع ریاضی $\{X^2, X^3, \sqrt[3]{X}, \sqrt{X}, \text{Log}(X)\}$ استفاده شد. گام سوم، شامل انتخاب ساختار و معماری کروموزوم ها می باشد. گام چهارم انتخاب تابع پیوندی است که در این مطالعه عمل جمع برای برای ایجاد پیوند بین زیرشاخه ها مورد استفاده قرار گرفت. در نهایت، در گام پنجم عملگرهای ژنتیک و نرخ هریک از آنها انتخاب می شود.

ماشین های بردار پشتیبان

در یک مدل رگرسیونی SVM لازم است وابستگی تابعی متغیر وابسته y به مجموعه ای از متغیرهای مستقل x تخمین زده شود. فرض بر این است که مانند دیگر مسائل رگرسیونی، رابطه بین متغیرهای وابسته و مستقل توسط یک تابع معین f به علاوه یک مقدار اضافی نویز^۲ مشخص می شود.

$$y = f(x) + \text{Noise} \quad (1)$$

بنابراین موضوع اصلی، پیدا کردن فرم تابع f است که بتواند به صورت صحیح، موارد جدیدی را که SVM تاکنون تجربه نکرده است پیش بینی کند. این تابع به وسیله آموزش مدل SVM بر روی یک مجموعه داده به عنوان مجموعه آموزش که شامل فرآیندی به منظور بهینه سازی دائمی تابع خطا است، قابل دسترسی است. بر مبنای تعریف این تابع خطا، دو نمونه از مدل های SVM شناخته شده است که عبارتند از الف) مدل های رگرسیونی SVM نوع اول که مدل های

3 - Linear kernel

4 - Polynomial kernel

5 - Hyperbolic tangent kernel

6 - Radial Base Function kernel

1 - Boolean functions

2 - Noise

SVM است (۱۹).

$$f(x_i) = \sum_{i=1}^n (-\partial_i - \partial_i^*) K(x_i, x_j) + b \quad (۴)$$

معیارهای ارزیابی مدل

در این تحقیق، برای ارزیابی مدل‌های مورد نظر از معیارهای ضریب همبستگی و جذر میانگین مربعات خطا استفاده می‌شود:

$$R = \left(1 - \frac{\sum_{i=1}^n (Q_i - \hat{Q}_i)^2}{\sum_{i=1}^n (Q_i - \bar{Q})^2} \right)^{0.5} \quad (۵)$$

$$RMSE = \left(\frac{\sum_{i=1}^n (Q_i - \hat{Q}_i)^2}{n} \right)^{0.5} \quad (۶)$$

که در روابط فوق Q_i مقدار مشاهده شده در گام زمانی i ام، \hat{Q}_i مقدار محاسبه شده در همان زمان، n تعداد داده‌ها و \bar{Q} میانگین مقادیر مشاهداتی می‌باشد. هر چه مقدار ضریب همبستگی بالاتر و RMSE کمتر باشد، نتایج دقیق‌تر و قابل اعتمادتر خواهند بود.

نتایج

نتایج مدل‌سازی برنامه‌ریزی ژنتیک

در مطالعه حاضر برای مدل‌سازی جریان روزانه رودخانه باراندوزچای با استفاده از برنامه‌ریزی ژنتیک، داده‌های ۴ سال (از یک مهر ۱۳۸۵ تا سی و یک شهریور ۱۳۸۸) برای آموزش و یک سال (از

یک مهر ۱۳۸۸ تا سی و یک شهریور ۱۳۸۹) به عنوان داده‌های تست انتخاب شدند. به طور کلی ۸۰ درصد داده‌ها برای آموزش و ۲۰ درصد نیز برای تست در نظر گرفته شد. در برنامه‌ریزی ژنتیک انتخاب جمعیت‌های اولیه که همان الگوی‌های ورودی می‌باشد از اهمیت بالایی برخوردار است. با توجه به اینکه در این مطالعه توالی دبی روزهای قبل در پیش‌بینی دبی امروز مد نظر بوده، صرفاً از داده‌های دبی جریان با توالی برگشتی تا ۵ روز قبل به عنوان داده‌های آموزشی، به صورت ترکیب‌های مختلف مطابق با جدول ۳ استفاده شده است. نکته مهم دیگر در برنامه‌ریزی ژنتیک، انتخاب عملگرهای مدل برای انجام محاسبات می‌باشد. در این مطالعه عملگر ترکیب به دلیل دارا بودن دقت بالا، برای مدل‌سازی جریان روزانه انتخاب شد.

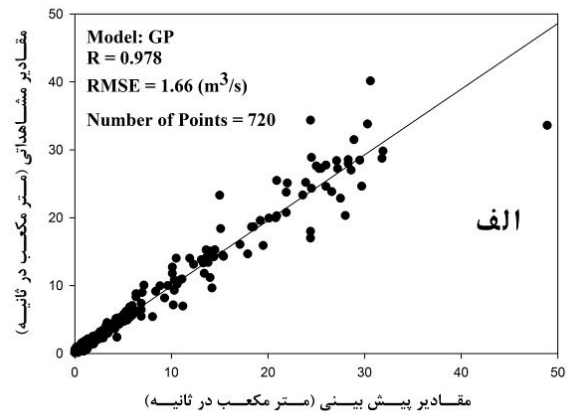
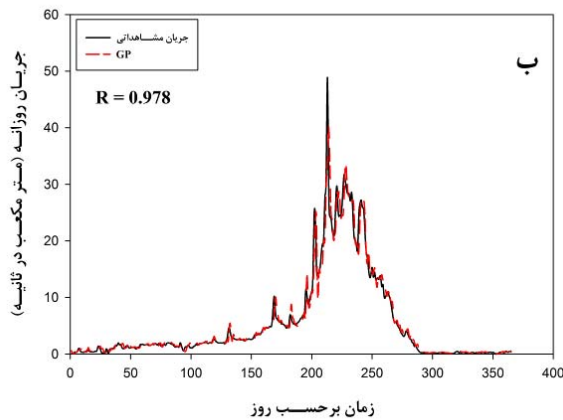
در جدول ۴ نتایج ارزیابی الگوهای مختلف ورودی برنامه‌ریزی ژنتیک برای مدل‌سازی جریان روزانه رودخانه در مراحل آموزش و صحت‌سنجی را نشان می‌دهد. با توجه به این جدول می‌توان نتیجه گرفت که مدل شامل دبی‌های یک، دو و سه روز قبل با $R=0.978$ و $RMSE=1/66 (m^3/s)$ دارای بیشترین دقت در مرحله صحت‌سنجی بوده و به عنوان بهترین الگو برای مدل برنامه‌ریزی ژنتیک انتخاب گردید. شکل ۳ نمودار مقادیر مشاهداتی و مقادیر پیش‌بینی شده حاصل از بهترین مدل برنامه‌ریزی ژنتیک (الگوی ۳) برای سری روزانه جریان رودخانه باراندوزچای، در مرحله صحت‌سنجی را نشان می‌دهد. با توجه به این شکل‌ها می‌توان نتیجه گرفت که برنامه‌ریزی ژنتیک در بهترین حالت خود به خوبی توانسته دبی جریان روزانه رودخانه مورد مطالعه را پیش‌بینی کند. همانگونه که در شکل ۳ مشاهده می‌شود، دقت نتایج مدل برنامه‌ریزی ژنتیک در پیش‌بینی دبی‌های کمتر بیش از دبی‌های بزرگتر می‌باشد.

جدول ۳- الگوهای ورودی مورد استفاده در مقیاس زمانی روزانه رودخانه باراندوزچای

ردیف	الگوی ورودی مدل در مقیاس روزانه
۱	$Q(t)=f\{Q(t-1)\}$
۲	$Q(t)=f\{Q(t-1), Q(t-2)\}$
۳	$Q(t)=f\{Q(t-1), Q(t-2), Q(t-3)\}$
۴	$Q(t)=f\{Q(t-1), Q(t-2), Q(t-3), Q(t-4)\}$
۵	$Q(t)=f\{Q(t-1), Q(t-2), Q(t-3), Q(t-4), Q(t-5)\}$

جدول ۴- تحلیل‌های آماری نتایج برنامه‌ریزی ژنتیک برای الگوهای مختلف ورودی جریان روزانه رودخانه باراندوزچای

شماره الگو	آموزش		صحت‌سنجی	
	R	RMSE (m^3/s)	R	RMSE (m^3/s)
۱	۰/۹۰۴	۳/۱۸	۰/۹۷۴	۱/۹۳
۲	۰/۹۳۴	۲/۶۶	۰/۹۷۵	۱/۷۸
۳	۰/۹۳۹	۲/۵۶	۰/۹۷۸	۱/۶۶
۴	۰/۹۳۰	۲/۷۳	۰/۹۷۴	۱/۸۲
۵	۰/۹۰۹	۲/۱۶	۰/۹۷۳	۱/۸۸



شکل ۳- نمودارهای الف) پراکندگی و ب) ترسیمی مقادیر مشاهداتی و مقادیر پیش بینی شده حاصل از بهترین مدل برنامه ریزی ژنتیک در مرحله صحت سنجی

چن و یو (۹) از برنامه اصلاح شده الگوریتم جستجوی شبکه که به نام الگوریتم جستجوی شبکه دو مرحله‌ای^۲ معروف است به همراه اعتبارسنجی متقاطع^۳ استفاده شد. برای این منظور ابتدا با انتخاب شبکه‌هایی با ابعاد بزرگ محدوده مشخصه‌های ϵ و C به ازای مقدار ثابت مشخصه γ تعیین شد. سپس با مشخص شدن محدوده مذکور و تقسیم آن به شبکه‌هایی با ابعاد ریزتر مقادیر دقیق دو مشخصه ϵ و C مشخص شدند. روند مذکور برای دیگر مقادیر γ نیز تکرار شد و بدین طریق مدل‌های متفاوتی با تغییر در مقدار γ حاصل شدند. حال می‌توان از بین مدل‌های توسعه داده شده مدل با کمترین خطا را تعیین کرده و مشخصه‌های آن را به عنوان مقادیر بهینه ϵ ، C و γ انتخاب نمود.

در تحقیق حاضر از مدل SVM نیز با همان الگوهای ورودی برنامه ریزی ژنتیک برای مدل بندی جریان رودخانه استفاده شد. برای هر الگوی ورودی با تغییر ϵ ، C و γ شبکه‌های مختلفی ساخته و آموزش داده شد و در نهایت ساختاری که دارای کمترین خطا بود، به عنوان مناسب‌ترین الگو انتخاب شد. جدول ۵ شاخص‌های آماری مربوط به نتایج حاصل از کاربرد مدل SVM و مناسب‌ترین مقادیر ϵ ، C و γ را برای هر الگوی ورودی نشان می‌دهد. از نتایج این جدول چنین بر می‌آید که دقت مدل سازی رفته رفته تا سه تاخیر زمانی دبی رو به بهبود بوده و بعد از آن کاهش پیدا می‌کند. از مقایسه جداول ۴ و ۵ می‌توان نتیجه گرفت که برنامه ریزی ژنتیک از دقت بهتری نسبت به مدل SVM برخوردار بوده و در تخمین دبی پیک نیز عملکرد بهتری داشته است. همچنین مدل SVM در تخمین مقادیر کمینه پایانی نیز دچار بیش برآزش شده است (شکل ۵).

یکی از قابلیت‌های مهم روش برنامه ریزی ژنتیک، ارائه رابطه ریاضی برای متغیرهای هیدرولوژیکی مورد بررسی می‌باشد. در این مطالعه نیز رابطه ریاضی به دست آمده از برنامه ریزی ژنتیک به همراه ساختار آن ارائه شده است. شکل ۴ ساختار درختی هریک از ژن‌ها را نشان می‌دهد. در این ساختار C_0 و C_1 ضرایب برنامه ریزی ژنتیک و d_1 ، d_2 و d_0 به ترتیب معادل دبی یک، دو و سه روز قبل می‌باشند. با استفاده از این ساختار درختی می‌توان شکل ریاضی جریان روزانه رودخانه باراندوزچای را به صورت رابطه ۷ ارائه نمود.

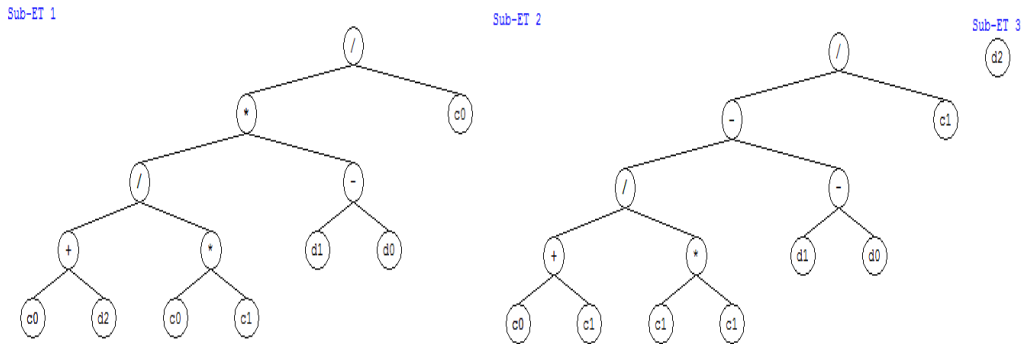
$$Q_t = 0.35 + 0.425Q_{t-2} - 0.425Q_{t-3} - 0.026Q_{t-1}Q_{t-2} + 0.026Q_{t-2}Q_{t-3} + Q_{t-1} \quad (7)$$

نتایج مدل سازی SVM

در این مرحله به منظور کاهش دامنه تغییرات داده‌های جریان روزانه رودخانه باراندوزچای و همسان سازی اطلاعات ورودی و خروجی، داده‌ها استانداردسازی شد. در مرحله بعد مقادیر بهینه مشخصه‌های مدل SVM شامل ϵ و C تعیین می‌گردد. همچنین در این مطالعه تابع کرنل مورد استفاده، تابع RBF انتخاب شد چرا که از دقت بهتری در برآورد جریان روزانه رودخانه باراندوزچای برخوردار بود. در این تابع نیز مشخصه γ بایستی تعیین گردد. بنابراین در حالت کلی برای پیش بینی جریان روزانه رودخانه باراندوزچای توسط مدل SVM رگرسیون، لازم است که مقادیر بهینه سه مشخصه مذکور به دست آید که بدین منظور دو مشخصه ϵ و C توسط الگوریتم بهینه سازی جستجوی شبکه^۱ و متغیر γ نیز به صورت آزمون و خطا محاسبه شد. البته قابل ذکر است که الگوریتم بهینه سازی جستجوی شبکه بسیار کند عمل می‌کند و زمان محاسباتی زیادی را به خود اختصاص می‌دهد. برای حل این مشکل در تحقیق مذکور طبق توصیه

2 - Two-Steps Grid Search
3 - Cross-Validation

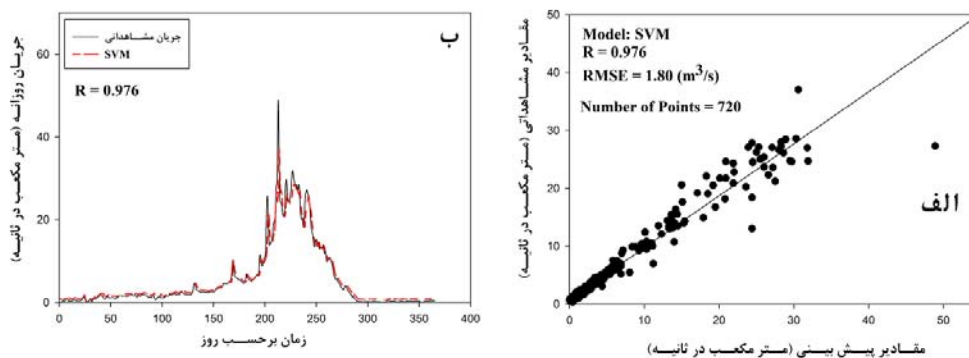
1 - Grid Search



شکل ۴- ساختار درختی برنامه‌ریزی ژنتیک برای بهترین مدل جریان روزانه رودخانه باراندوزچای

جدول ۵- تحلیل‌های آماری نتایج SVM برای الگوهای مختلف ورودی جریان روزانه رودخانه باراندوزچای

شماره الگو	آموزش		صحت‌سنجی		متغیرهای مدل SVM		
	R	RMSE (m ³ /s)	R	RMSE (m ³ /s)	C	ε	γ
۱	۰/۹۳۵	۲/۸۱	۰/۹۷۵	۲/۰۵	۹	۰/۲۰	۱/۵۰۰
۲	۰/۹۳۵	۲/۷۵	۰/۹۷۴	۱/۹۹	۱۰	۰/۳۰	۰/۸۵۹
۳	۰/۹۳۶	۲/۷۷	۰/۹۷۶	۱/۸۰	۱۰	۰/۱۰	۰/۳۲۹
۴	۰/۹۳۴	۲/۹۹	۰/۹۷۵	۲/۳۶	۱۰	۰/۱۰	۰/۳۷۰
۵	۰/۹۳۳	۲/۶۸	۰/۹۷۴	۲/۴۱	۱۰	۰/۴۰	۰/۲۱۰



شکل ۵- نمودارهای الف) پراکندگی و ب) ترسیمی مقادیر مشاهداتی و مقادیر پیش‌بینی شده حاصل از بهترین مدل SVM در مرحله صحت‌سنجی

نتیجه‌گیری

میانگین مربعات مجذور خطا برای این الگو به ترتیب ۰/۹۷۸ و $1/66 (m^3/s)$ محاسبه گردید.

مقادیر بهینه مشخصه‌های مدل SVM یعنی ϵ و C با استفاده از الگوریتم بهینه‌سازی جستجوی شبکه و مشخصه γ با استفاده از سعی و خطا به ترتیب معادل ۰/۱۰، ۱۰ و ۰/۳۲۹ به دست آمد. ارزیابی عملکرد SVM بر مبنای دو آماره R و RMSE نشان می‌دهد که دقت مدل‌سازی تا تأخیر سوم رو به بهبود بوده و از آن به بعد کاهش می‌یابد. به طوری که ضریب همبستگی و میانگین مربعات مجذور خطا در بهترین حالت مدل SVM به ترتیب معادل ۰/۹۷۶ و

تعیین جریان رودخانه‌ها یکی از مولفه‌های مهم و تاثیرگذار در مدیریت منابع آب حوضه آبریز می‌باشد. در این مطالعه جریان روزانه رودخانه باراندوزچای در دوره آماری ۱۳۸۵ تا ۱۳۸۹ با استفاده از روش برنامه‌ریزی ژنتیک و ماشین بردار پشتیبان مدل‌سازی شد. به طور کلی نتایج حاصل از این مطالعه به شرح زیر ارائه می‌شود:
برای سری روزانه بهترین عملکرد برنامه‌ریزی ژنتیک با سه تأخیر زمانی در ورودی‌های مدل به دست آمد و ضریب همبستگی و

$$(m^3/s) / 1/80 \text{ به دست آمد.}$$

مدل رگرسیون از خود نشان داد. فریبودنام و همکاران (۲) و داننده مهر و همکاران (۱) و قبادیان و همکاران (۳) نیز در تحقیقات خود نشان دادند که برنامه ریزی ژنتیک از دقت بسیار بالایی در مدل سازی فرایندهای مختلف برخوردار است و نتیجه این تحقیق نیز برتری روش GP را نسبت به مدل SVM نشان می دهد.

نتایج این مطالعه نشان داد که گرچه دقت روش برنامه ریزی ژنتیک نسبت به روش ماشین بردار پشتیبان اندکی بیشتر بود، اما روش SVM به مراتب ساده تر از روش GP می باشد و می تواند به عنوان یک روش کاربردی برای پیش بینی جریان روزانه رودخانه ها بکار گرفته شود.

نتایج مطالعه حاضر در مجموع روش برنامه ریزی ژنتیک را به عنوان یک روش صریح و دقیق برای پیش بینی جریان رودخانه ها پیشنهاد می کند. این پیشنهاد با نتایج حاصل از مطالعات گون (۱۲) که به پیش بینی جریان روزانه رودخانه به کمک برنامه ریزی ژنتیک پرداخته است، مطابقت کامل دارد. خو و همکاران (۱۴) از برنامه ریزی ژنتیک برای پیش بینی رواناب بهره برده و نتایج حاصل را با مقادیر مشاهداتی و نیز مقادیر محاسبه شده توسط روش های کلاسیک مقایسه نمودند. نتایج بدست آمده حاکی از دقت بالای برنامه ریزی ژنتیک بود. در تحقیقی که آیتک و کیشی (۸) در مدل سازی رسوب معلق انجام دادند، برنامه ریزی ژنتیک دقت فوق العاده بالایی نسبت به

منابع

- ۱- داننده مهر ع. و مجدزاده طباطبائی م.ر. ۱۳۸۹. بررسی تأثیر توالی دبی روزانه در پیش بینی جریان رودخانه ها با استفاده از برنامه ریزی ژنتیک. - نشریه آب و خاک دانشگاه فردوسی مشهد. ۲۴(۲): ۳۲۵-۳۳۳.
- ۲- فریبودنام ن.، قربانی م.ع. و اعلمی م.ت. ۱۳۸۸. پیش بینی جریان رودخانه با استفاده از برنامه ریزی ژنتیک (مطالعه موردی: رودخانه ليقوان). مجله دانش آب و خاک (دانشگاه تبریز). ۱۹(۴): ۱۲۳-۱۰۷.
- ۳- قبادیان ر.، قربانی م.ع. و خلج م. ۱۳۹۲. بررسی عملکرد روش برنامه ریزی بیان ژن در روندیابی سیلاب رودخانه زنگمار در مقایسه با روش موج دینامیکی. نشریه آب و خاک. ۲۷(۳): ۶۰۲-۵۹۲.
- ۴- قربانی م.ع.، شیرازی ج. و کاظمی ه. ۱۳۸۹. تخمین بیشینه، متوسط و کمینه دمای شهر تبریز با استفاده از روش های هوش مصنوعی. مجله دانش آب و خاک (دانشگاه تبریز). ۲۰(۱): ۱۰۴-۸۷.
- ۵- نیک بخت شهپازی ع.ر. ۱۳۸۸. کاربرد ماشین بردار پشتیبان در پیش بینی جریان رودخانه. هشتمین کنفرانس هیدرولیک ایران. دانشکده فنی دانشگاه تهران.
- 6- Adamowski J., and Prasher S.O. 2012. Comparison of machine learning methods for runoff forecasting in mountainous watersheds with limited data. *Journal of Water and Land Development*, 17(7-8): 89-97 .
- 7- Alvisi S., Mascellani G., Franchini M., and Bardossy A. 2005. Water level forecasting through fuzzy logic and artificial neural network approaches. *J Hydrol Earth Sys Sci.*, 2:1107-1145.
- 8- Aytak A., and Kisi O. 2008. A genetic programming approach to suspended sediment modeling. *Journal of Hydrology* 351: 288-298.
- 9- Chen S.T., and P.S. Yu .2007. Real-time probabilistic forecasting of flood stages. *Journal of Hydrology*, 340: 63-77.
- 10- Ferreira C. 2001. Gene expression programming: a new adaptive algorithm for solving problems. *Complex Syst* 13: 87-129.
- 11- Ghorbani M.A., Kisi O. and Aalinezhad M. 2010. A probe into the chaotic nature of daily streamflow time series by correlation dimension and largest Lyapunov methods *Applied Mathematical Modelling*, 34: 4050-4057
- 12- Guven A. 2009. Linear genetic programming for time-series modeling of daily flow rate. *J. Earth Syst. Sci.*, 118(2):157-173.
- 13- Hamel L. 2009. *Knowledge Discovery with Support Vector Machines*, Hoboken, N.J. John Wiley.
- 14- Khu S.T., Liong S.Y., Babovic V., Madsen H., and Muttil N. 2001. Genetic programming and its application in real-time runoff forecasting. *Journal of American Water Resources Association*, 37 (2): 439-451.
- 15- Koza J.R. 1992. *Genetic Programming: on the programming of computers by means of natural selection*. Cambridge, MA: MIT Press.
- 16- Pai P.F., and Hong W.C. 2007. A recurrent support vector regression model in rainfall forecasting. *Hydrological Process*, 21:819-827.
- 17- Vapnik V.N. 1998. *Statistical Learning Theory*. Wiley, New York.
- 18- Yu P.S., Chen S.T., and Chang I.F. 2005. Flood stage forecasting using support vector machines. *Geophysical Research Abstracts*, Vol.7,04176.

- 19- Yu P.S, Chen S.T., and Chang I.F. 2006. Support vector regression for real-time flood stage forecasting. *Hydrology*, 328: 704-716.
- 20- Zahiri A., and Azamathulla H.Md. 2014. Comparison between linear genetic programming and M5 tree models to predict flow discharge in compound channels. *Neural Comput & Applic*, 24:413-420.



Comparison between Genetic Programming and Support Vector Machine Methods for Daily River Flow Forecasting (Case Study: Barandoozchay River)

F. Ahmadi^{1*} - F. Radmanesh² - R. Mirabbasi Najaf Abadi³

Received: 17-02-2014

Accepted: 01-09-2014

Abstract

Accurate estimation of river flow can have a significant importance in water resources management. In this study, Genetic programming (GP) and Support Vector Machine (SVM) methods were used to forecast daily discharge of Barandoozchay River. The daily discharge data of Barandoozchay River measured at the Dizaj hydrometric station during 2007 to 2011 was used for modeling, which 80% of the data used for training and remaining 20% used for testing of models. The results showed that in the both of considered methods, the models including discharges of one, two and three days ago had higher accuracy in verification step and the accuracy of models decreased with increasing discharge values. Comparing the performance of GP and SVM methods indicated that, however the accuracy of the GP method with the $R=0.978$ and $RMSE=1.66$ (m^3/s) was slightly more than SVM method with $R=0.976$ and $RMSE=1.80$ (m^3/s), but the SVM is easier than GP method. Thus, the SVM method can be used as an alternative method in forecasting daily river discharge.

Keywords: Genetic programming (GP), Support Vector Machine (SVM), Daily discharge forecasting, Barandoozchay River

1, 2 - PhD Candidate and Associate Professor of Water Engineering Resources Department, Faculty of Water Science, Shahid Chamran University, Ahvaz, Iran

(* - Corresponding Author Email: Farshad.paper@yahoo.com)

3 - Assistant Professor of Water Engineering Department, Sharekord University, Sharekord, Iran