

شبیه‌سازی خودمبسته جریان حوضه آبریز زرينه‌رود با استفاده از روش تجزیه پروکراستس و مدل‌های شبکه عصبی مصنوعی و ماشین بردار پشتیبان

بهروز سبجانی^{۱*}، محمد عیسی‌زاده^۲ و منیر شیرزاد^۳

^(۱) دانشیار؛ گروه آب و هواشناسی؛ دانشکده انسانی؛ دانشگاه محقق اردبیلی؛ اردبیل؛ ایران

*نویسنده مسئول مکاتبات: Sobhani@uma.ac.ir

^(۲) دانشجوی دکتری مهندسی منابع آب؛ گروه مهندسی آب؛ دانشکده کشاورزی؛ دانشگاه تبریز؛ تبریز؛ ایران

^(۳) دانشجوی کارشناسی ارشد سنجش از دور و سیستم اطلاعات جغرافیایی؛ گروه سنجش از دور و سیستم اطلاعات جغرافیایی؛ دانشکده جغرافیا و برنامه‌ریزی؛ دانشگاه تبریز؛ تبریز؛ ایران

تاریخ دریافت: ۱۳۹۵/۰۷/۲۵ تاریخ پذیرش: ۱۳۹۶/۰۳/۰۳

چکیده

پیش‌بینی جریان رودخانه‌ها در حوضه‌های آبریز نقش مهمی در بهره‌برداری و مدیریت صحیح منابع آبی دارد. تعیین نوع و تعداد ورودی مدل‌های تخمین‌گر، یکی از مهم‌ترین مراحل در پیش‌بینی جریان رودخانه‌ها می‌باشد. بنابراین از روش تجزیه پروکراستس (PA) برای تعیین تعداد ورودی‌های موثر استفاده شده است. در این تحقیق پیش‌بینی جریان با استفاده از داده‌های جریان ماهانه ایستگاه‌های آب‌سنجی صفاخانه و سننه انجام گرفته است. مدل شبکه عصبی مصنوعی (MLP) و مدل ماشین بردار پشتیبان (SVM) برای پیش‌بینی جریان مورد استفاده قرار گرفته‌اند. بهترین تخمین جریان با استفاده از مدل‌های MLP و PA-MLP در ایستگاه آب‌سنجی صفاخانه به ترتیب با RMSE برابر با $5/68$ (m^3/s) و $4/85$ (m^3/s) و CC برابر با $0/73$ و $0/78$ و در ایستگاه آب‌سنجی سننه به ترتیب با RMSE برابر با $6/44$ (m^3/s) و $6/36$ (m^3/s) و CC برابر با $0/78$ و $0/79$ انجام شده است. مدل PA-SVM نیز به ترتیب با RMSE و CC برابر با $5/45$ (m^3/s) و $0/73$ در دوره صحت‌سنجی نتایج بهتری را نسبت به مدل SVM در تخمین جریان ایستگاه آب‌سنجی صفاخانه داشته است. همچنین نتایج نشان داد که مدل‌های SVM و PA-SVM جریان ایستگاه سننه را با RMSE به ترتیب برابر با $6/85$ (m^3/s) و $7/03$ (m^3/s) تخمین زده‌اند. در حالت کلی نتایج نشان داد که روش تجزیه پروکراستس می‌تواند به‌عنوان یکی از روش‌های کارآمد و مناسب جهت تعیین تعداد ورودی موثر مورد استفاده قرار گیرد. مقایسه نتایج مدل‌های MLP و SVM نیز نشان داد که مدل MLP از دقت بیشتری نسبت به مدل SVM برخوردار است.

کلید واژه‌ها: پیش‌بینی جریان؛ تجزیه پروکراستس؛ شبکه عصبی مصنوعی؛ ماشین بردار پشتیبان

مقدمه

قرار گرفته است، پیش‌بینی جریان رودخانه‌های این حوضه از اهمیت خاصی برخوردار می‌باشند. مدل‌های مختلفی در تخمین جریان حوضه‌های آبخیز مورد استفاده قرار می‌گیرند که کارایی و همچنین موارد استفاده از آنها براساس ساختار حوضه، داده‌های در دسترس و مورد نیاز و همچنین هدف شبیه‌سازی از همدیگر متفاوت است. انتخاب مدلی که بتواند میزان جریان رودخانه را با دقت بالا و خطای کم مدل‌سازی نماید کمک شایانی در مدیریت

پیش‌بینی جریان رودخانه‌ها در تصمیم‌گیری و سیاست‌گذاری‌های حوضه‌های آبریز از اهمیت ویژه‌ای برخوردار است. این موضوع در مدیریت بهره‌برداری از آب مخازن سدها، انرژی برق آبی و ... به‌طور کامل مورد استفاده قرار می‌گیرد. با توجه به اینکه حوضه زرينه‌رود بیشترین جریان ورودی به دریاچه ارومیه را داراست و با توجه به اینکه سد بزرگ بوکان نیز بر روی این رودخانه

نتایج حاکی از دقت بیشتر مدل شبکه عصبی مصنوعی نسبت به دو مدل دیگر در پیش‌بینی جریان ماهانه بوده است. ازانی (۱۳۹۳) نیز براساس مدل‌های شبکه عصبی مصنوعی، ماشین بردار پشتیبان و موجک این دو مدل، فرآیند بارش-رواناب حوضه صوفی‌چای را مورد شبیه‌سازی قرار داده‌اند. نتایج حاکی از دقت بیشتر مدل موجک-ماشین بردار پشتیبان در حالت کلی و دقت بیشتر مدل موجک-عصبی در تخمین مقادیر حدی می‌باشد. عیسی‌زاده (۱۳۹۴) به‌ازای شصت ترکیب مختلف جریان سه رودخانه حوضه دریاچه ارومیه را مورد تخمین قرار داده‌اند. نتایج این تحقیق نشان داد که مدل شبکه عصبی مصنوعی دقت بیشتری نسبت به مدل ماشین بردار پشتیبان داشته است. همچنین این تحقیق نشان داد که در تعیین تعداد ورودی‌ها نبایستی تنها به چند تاخیر اول اکتفا نمود و تعداد ورودی‌های بهینه بایستی براساس سعی و خطا بدست آید.

Dinpashoh و همکاران (۲۰۰۴) جهت پهنه‌بندی اقلیم بارش کشور ایران از ۵۷ متغیر مرتبط با بارش بدست آمده از ۷۷ ایستگاه هواشناسی استفاده کرده‌اند. سپس با استفاده از روش پروکراستس تعداد ۱۲ متغیر مهم از بین متغیرهای کنديدا انتخاب گردیده است. آنگاه با اعمال تجزیه به عامل و روش تجزیه خوشه‌ای وارد (Ward)، هفت گروه مشابه اقلیم بارش ایران در نظر گرفته شد. نتایج آزمون همگنی نشان داد که شش ناحیه کاملاً همگن و یک ناحیه غیرهمگن می‌باشد. Asefa و همکاران (۲۰۰۵) مدل SVM با تابع کرنل خطی را برای پیش‌بینی جریان‌های فصلی و ساعتی در حوضه رودخانه سویتر در ایالات متحده بکار گرفتند. آنها در این تحقیق با توجه به حجم جریان معادل برف و همچنین حجم جریان در دوره‌های پیشین، میزان حجم جریان در دوره‌های ۶ ماهه و ۲۴ ساعته را پیش‌بینی نمودند که نتایج مطالعه ایشان از دقت مطلوبی برخوردار بود.

Misra و همکاران (۲۰۰۹) عملکرد مدل‌های SVM و ANN در پیش‌بینی رواناب و بار رسوب حوضه آبریز

صحیح طرح‌های منابع آبی و افزایش عملکرد آنها دارد. در دهه‌های اخیر مدل‌های داده‌کاوی کاربرد گسترده‌ای در زمینه شبیه‌سازی و پیش‌بینی جریان رودخانه‌ها پیدا کرده‌اند (Samsudin et al., 2011). این مدل‌های متکی به داده، ساختار ریاضی مشخصی داشته و بدون اینکه فرآیند فیزیکی حاکم بر پدیده را شبیه‌سازی نمایند، توانایی تشخیص روابط پیچیده غیرخطی بین داده‌های ورودی و خروجی را دارا هستند (Nayak et al., 2004). شبکه عصبی مصنوعی^۱ (ANN) و ماشین بردار پشتیبان^۲ (SVM) از مهم‌ترین مدل‌های داده‌کاوی هستند که در دهه‌های اخیر جهت تخمین جریان رودخانه‌ها مورد توجه محققین قرار گرفته‌اند.

در مدل‌های داده‌کاوی به‌دلیل ماهیتی که دارند ساختار مدل، نوع و ترکیب داده‌های بکار گرفته شده در شبیه‌سازی، تأثیر به‌سزایی در عملکرد آنها دارد (نوری و همکاران، ۱۳۸۹). بنابراین تعیین تعداد و نوع ورودی مدل‌ها نقش مهمی در افزایش دقت و کاهش خطای مدل‌ها دارد. روش‌های خطی و غیرخطی زیادی جهت تعیین ورودی‌ها مورد استفاده قرار گرفته است. در این مطالعه از روش پیشنهادی Krzanowski (۱۹۸۷) به نام روش تجزیه پروکراستس (Procrustes Analysis) استفاده شده است. این روش از بین متغیرهای ورودی، آن مجموعه‌ای که بیشترین اطلاعات کل مجموعه داده‌ها را دارا باشد، انتخاب می‌کند. تاکنون تحقیقات زیادی در مورد ارزیابی و همچنین مقایسه عملکرد مدل‌های تخمین‌گر در تخمین جریان رودخانه انجام گرفته است.

نوری و همکاران (۱۳۸۹) تأثیر انتخاب متغیرهای ورودی با استفاده از تجزیه مؤلفه اصلی بر عملکرد مدل SVM در پیش‌بینی دبی ماهانه رودخانه صوفی‌چای را مورد بررسی قرار دادند. بنی‌حیب و همکاران (۱۳۹۰) عملکرد سه مدل آرما، آرما و شبکه عصبی مصنوعی را در پیش‌بینی ماهانه جریان ورودی به مخزن سد دز مورد مقایسه قرار دادند که

¹ Artificial Neural Network

² Support Vector Machine

همچنین Kalteh (۲۰۱۳) ترکیب مدل‌های رگرسیون بردار پشتیبان (Support Vector Regression (SVR)) و شبکه عصبی مصنوعی با تبدیل موجک را جهت پیش‌بینی جریان ماهانه بکار گرفتند و به این نتیجه رسیدند که هر دو مدل رگرسیون بردار پشتیبان و شبکه عصبی مصنوعی در ترکیب با تبدیل موجک کارایی بهتری نسبت به مدل‌های رگرسیون بردار پشتیبان و شبکه عصبی مصنوعی تنها دارند. Terzi و Ergin (۲۰۱۴) مدل‌های AR، برنامه‌ریزی بیان ژن، شبکه عصبی پایه شعاعی^۱ (RBF)، شبکه عصبی پیشرو^۲ (FFNN) و نروفازی را جهت پیش‌بینی رواناب ماهانه رودخانه کیزیلیرماک (Kizilirmak) در ترکیه مورد استفاده قرار دادند. نتایج نشان داد که مدل AR(۲) از بین مدل‌های بکار گرفته شده در تحقیق عملکرد بهتری را در پیش‌بینی جریان داشت. در تحقیقی دیگر Awchi (۲۰۱۴) دقت مدل‌های FFNN، RBF، شبکه‌های عصبی رگرسیونی و رگرسیون چندگانه را در پیش‌بینی جریان ماهانه رودخانه‌های زاب بالادست و پایین دست در شمال عراق مورد مقایسه قرار دادند و به این نتیجه رسیدند که مدل‌های شبکه عصبی مصنوعی عملکرد بهتری از مدل رگرسیون چندگانه در پیش‌بینی جریان دارند.

Nam و همکاران (۲۰۱۵) آنالیز فرکانس منطقه‌ای حداکثر بارش سالانه را با استفاده از داده‌های ۶۷ ایستگاه هواشناسی کشور کره‌شمالی انجام داده‌اند. در این تحقیق با استفاده از روش پروکراستس از کل ۴۲ متغیر در نظر گرفته شده، تعداد ۳۳ متغیر به‌عنوان متغیرهای اساسی‌تر انتخاب گردیدند. سپس از این متغیرهای منتخب و روش‌های خوشه‌بندی وارد و فازی - C-means جهت تقسیم‌بندی ایستگاه‌ها استفاده گردیده است که این روش‌ها پنج ناحیه همگن را برای بارش کره‌جنوبی تعیین کرده‌اند. Ghorbani و همکاران (۲۰۱۶) با استفاده از مدل‌های ماشین بردار پشتیبان، شبکه عصبی مصنوعی پرسپترون چندلایه و پایه شعاعی جریان ماهانه رودخانه‌های حوضه

وامسادهارا (Vamsadhara) را واقع در جنوب کشور هند را مورد مقایسه قرار دادند. نتایج نشان داد مدل SVM هم در پیش‌بینی رواناب و هم بار رسوب نسبت به ANN دارای عملکرد بهتری می‌باشد. همچنین Liu (۲۰۱۱) در مطالعه خود کارایی مدل‌های رگرسیونی، ARIMA و ANN را در پیش‌بینی جریان روزانه رودخانه وایت‌کلی‌کریک نزدیک ایالت پنسیلوانیا مورد ارزیابی قرار داده و دقت آنها را با یکدیگر مقایسه کردند. نتایج تحقیق آنان نشان داد زمانی که داده‌ها از توزیع نرمال تبعیت کرده و دارای طول دوره آماری کافی باشند، مدل‌های خطی کمترین خطا را در پیش‌بینی جریان رودخانه دارا می‌باشند. در تحقیقی دیگر، Kisi و همکاران (۲۰۱۲) دقت مدل‌های نروفازی، شبکه عصبی مصنوعی و ماشین بردار پشتیبان را در پیش‌بینی رواناب روزانه در دو ایستگاه آب‌سنجی واقع در شمال غرب ترکیه مورد مقایسه قرار دادند. نتایج تحقیق آنان نشان داد که مدل‌های شبکه عصبی مصنوعی و نروفازی بترتیب بالاترین دقت را در پیش‌بینی رواناب روزانه در ایستگاه‌های اول و دوم دارا هستند. Adamowski (۲۰۱۳) مدل ماشین بردار پشتیبان و شبکه عصبی مصنوعی را برای شبیه‌سازی بارش-رواناب در یک حوضه کوهستانی با داده‌های محدود در یوتارانچال هندوستان به‌کار گرفت. نتایج بررسی وی نشان از توانایی مدل ماشین بردار پشتیبان با تابع کرنل پایه شعاعی در پیش‌بینی رواناب مستقیم، جریان پایه و جریان کل این منطقه کوهستانی و پر از تپه داشته است. Kakaei و همکاران (۲۰۱۳) سه مدل ماشین بردار پشتیبان و نروفازی (ANFIS) و مدل هیدرولوژیکی مفهومی Mike11/NA را در شبیه‌سازی جریان روزانه حوضه اسکندری به‌کار گرفتند. نتایج نشان داد که هر سه مدل دارای عملکرد مناسب در منطقه مورد مطالعه می‌باشند. در نهایت مدل ماشین بردار پشتیبان با تابع کرنل پایه شعاعی بالاترین ضریب همبستگی و کمترین میانگین مربعات خطا را در شبیه‌سازی جریان رودخانه نسبت به دو مدل دیگر از خود نشان داد.

¹ Radial Basis Function Network

² Feed-Forward Neural Networks

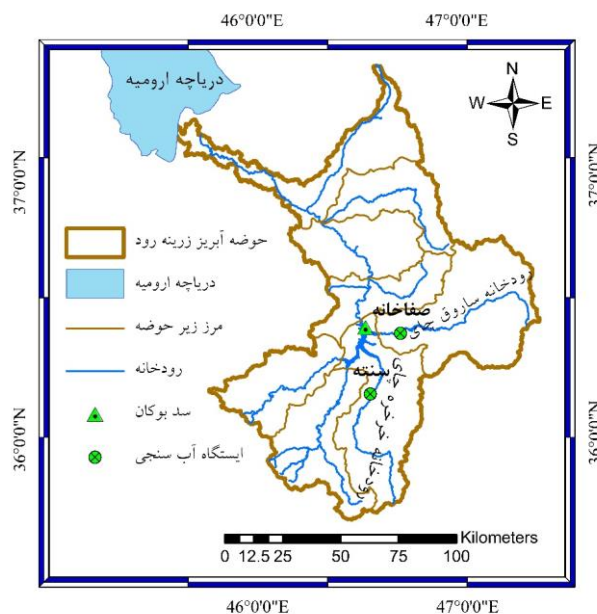
مواد و روش‌ها

داده‌ها و منطقه مورد مطالعه

در این تحقیق حوضه‌های آبریز خرخره‌چای و ساروق‌چای که از زیرحوضه‌های اصلی حوضه زرينه‌رود محسوب می‌شوند، به عنوان محدوده مورد مطالعه انتخاب گردیدند. این حوضه‌ها در در طول جغرافیایی $46^{\circ}21'$ تا $47^{\circ}20'$ شرقی و عرض جغرافیایی $35^{\circ}41'$ تا $36^{\circ}52'$ شمالی قراردارند (شکل ۱). این دو حوضه سهم قابل توجهی در حجم آب ورودی به سد بوکان و در نتیجه تأمین نیازهای آبی پایین دست این سد دارند. ایستگاه‌های سته و صفاخانه که به ترتیب در حوضه‌های خرخره‌چای و ساروق‌چای واقع شده‌اند از جمله ایستگاه‌های آب‌سنجی هستند که از طول دوره آماری قابل قبول برخوردار می‌باشند. مساحت حوضه بالادست هر یک از ایستگاه‌های آب‌سنجی مذکور به ترتیب بالغ بر ۱۲۳۲ و ۲۲۰۹ کیلومتر مربع می‌باشد. داده‌های مورد استفاده در این تحقیق شامل جریان ماهانه مشاهداتی در ایستگاه‌های آب‌سنجی سته و صفاخانه است که ۷۵ درصد داده‌های دوره آماری (۱۳۸۴-۱۳۶۷) به عنوان داده‌های واسنجی و ۲۵ درصد دیگر (۱۳۹۰-۱۳۸۵) به عنوان داده‌های صحت‌سنجی هر یک از مدل‌ها در نظر گرفته شده است.

دریاچه ارومیه را مورد پیش‌بینی قرار دادند. نتایج حاکی از برتری مدل‌های شبکه عصبی مصنوعی پرسپترون چندلایه و پایه شعاعی نسبت به مدل ماشین بردار پشتیبان بوده است. البته عدم قطعیت مدل ماشین بردار پشتیبان از دو مدل دیگر کمتر می‌باشد.

با توجه به آنچه آمد، تحقیق حاضر سعی دارد تا کارایی مدل‌های شبکه عصبی مصنوعی و ماشین بردار پشتیبان را در شبیه‌سازی خودهمبسته جریان ماهانه حوضه زرينه‌رود مورد ارزیابی قرار دهد و دقت این مدل‌ها را در این شبیه‌سازی‌ها نسبت به یکدیگر بررسی کند. لازم به ذکر است که روش‌های زیادی برای تعیین تعداد و نوع ورودی مدل‌ها وجود دارد، ولی نتایج هر یک از این روش‌ها متفاوت می‌باشد. بنابراین در این تحقیق سعی گردید علاوه بر روش‌های دیگر، برای اولین بار از روش تجزیه پروکراسس برای تعیین ورودی‌های مناسب جهت پیش‌بینی جریان خودهمبسته استفاده گردد. برای این منظور مدل شبکه عصبی پرسپترون چند لایه (MLP) و مدل ماشین بردار پشتیبان با سه تابع کرنل پایه شعاعی، چند جمله‌ای و کرنل خطی جهت شبیه‌سازی ماهانه رواناب مورد استفاده قرار گرفته‌اند.



شکل ۱. حوضه آبریز زرينه‌رود و محل قرارگیری ایستگاه‌های آب‌سنجی صفاخانه و سته

روش تجزیه به مولفه‌های اصلی (PCA)

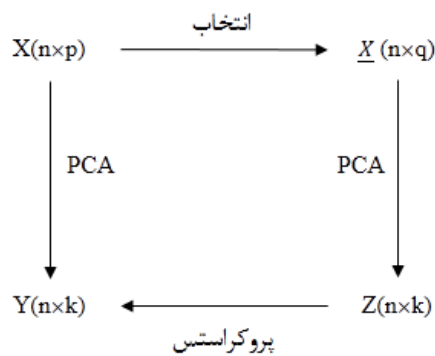
این روش شاید قدیمی‌ترین روش چند متغیره بوده و اولین بار در سال ۱۹۰۱ توسط پیرسون و سپس در سال ۱۹۳۳ توسط هتلینگ معرفی شد (Jolliffe, 1986). ایده اصلی PCA کاهش ابعاد داده‌هایی است که با هم همبستگی نسبی بالایی (از نظر قدر مطلق) دارند. کاهش ابعاد داده‌ها به کمک تبدیل این داده‌ها به متغیرهای جدید (PCها یا مولفه‌های مستقل از هم) آن‌چنان صورت می‌گیرد که اولاً چند مولفه اول تعیین کننده بخش زیادی از تغییرات کل داده‌ها باشند و ثانیاً تغییرات مولفه اول بیشتر از دوم، مولفه دوم بیشتر از سوم و ... باشد. محاسبه PCS برای هر متغیر مستلزم پیدا کردن مقادیر ویژه و بردارهای ویژه ماتریس پخشیدگی می‌باشد (Johnson and Hanson, 1995). اطلاعات بیشتر درباره این روش در مقاله Jolliffe (۱۹۸۶) وجود دارد.

روش تجزیه پروکراسستس (PA)

اگر بخواهیم سیمای کلی یک سری از داده‌های چندمتغیره را با یکی از روش‌های کاهش ابعاد داده‌ها، مانند PCA نشان دهیم لازم است تعداد مولفه‌های غالب (k)

مشخص شده و سری مقادیر مولفه‌های اصلی منتخب در مورد این مولفه‌ها محاسبه شوند. مقادیر مولفه‌ها ترکیب خطی از همه متغیرها (p) می‌باشند. اگر تعداد متغیرهای مشاهداتی (p) زیاد باشد، تعداد ضرایب نیز زیاد خواهد بود. بنابراین نام گذاری و تفسیر مولفه‌ها با مشکل روبه‌رو می‌گردد. یکی از راه‌های حل این مشکل، انتخاب تعداد معقولی متغیر (q) از بین همه متغیرها (p) می‌باشد، به طوریکه اطلاعات همه متغیرهای کانیدیدا به طور تقریبی در بطن متغیرهای انتخاب شده موجود باشد.

در این مطالعه از روش پیشنهادی Krzanowski (۱۹۸۷) استفاده شده است که دارای دو شرط می‌باشد. شرط اول: تعداد عناصر زیرمجموعه منتخب نسبت به مجموعه مرجع خیلی کم باشد ($q \ll p$) و شرط دوم: تعداد این عناصر برابر یا بیشتر از تعداد مولفه‌های اصلی انتخاب شده باشد ($q \geq k$). در این روش انتخاب متغیرها براساس معیاری موسوم به پروکراسستس که با " M^2 " نمایش داده می‌شود، انجام می‌گیرد. شکل ۲ چرخه کلی روش تجزیه پروکراسستس را نشان داده است.



شکل ۲. چرخه انتخاب متغیرها با روش پروکراسستس (Krzanowski, 1987)

در شکل ۲، x ماتریس کل داده‌های استاندارد شده و \bar{X} ماتریس داده‌های استاندارد شده به‌ازای حذف مرحله‌ای متغیرها می‌باشند. ماتریس Y که براساس همه متغیرهای کانیدیدا محاسبه می‌شود را اصطلاحاً آرایش واقعی و ماتریس Z که مبتنی بر متغیرهای منتخب می‌باشد را به‌عنوان آرایش تقریبی داده‌ها می‌نامند. تجزیه پروکراسستس

اختلاف مجموع مربعات نقاط متناظر این دو آرایش را ارزیابی می‌کند. هرچه متغیرهای منتخب دقیق‌تر انتخاب شوند، این دو آرایش مشابه هم بوده و تفاوت بین آرایش تقریبی و واقعی کمتر خواهد بود. حاصل نتایج تجزیه پروکراسستس محاسبه اختلاف بین مجموعه مربعات دو

می‌گردد. بهترین زیرمجموعه مرجع متغیرها چنان انتخاب می‌گردد که کمترین مقدار کمیت M^2 متعلق به آن زیرمجموعه باشد (Krzanowski, 1987). شکل ۳ روند کلی محاسبات روش تجزیه پروکراستس بکار رفته در این تحقیق را نشان داده است. جهت کسب اطلاعات بیشتر درباره روش پروکراستس به Krzanowski (۱۹۸۷) و دین‌پژوه (۱۳۸۲) مراجعه گردد. کلیه مراحل نشان داده شده در شکل ۳، در نرم افزار متلب کدنویسی شده است.

مدل ماشین بردار پشتیبان

اولین کاربرد این روش در مسائل آب توسط Dibike و همکاران (۲۰۰۱) با شبیه‌سازی بارش- رواناب ارائه شد.

آرایش تقریبی و واقعی می‌باشد که از معادله ۱ بدست می‌آید.

$$M^2 = \text{Trace} \{ Y Y' + Z Z' - 2 Z Q Y' \} \quad (1)$$

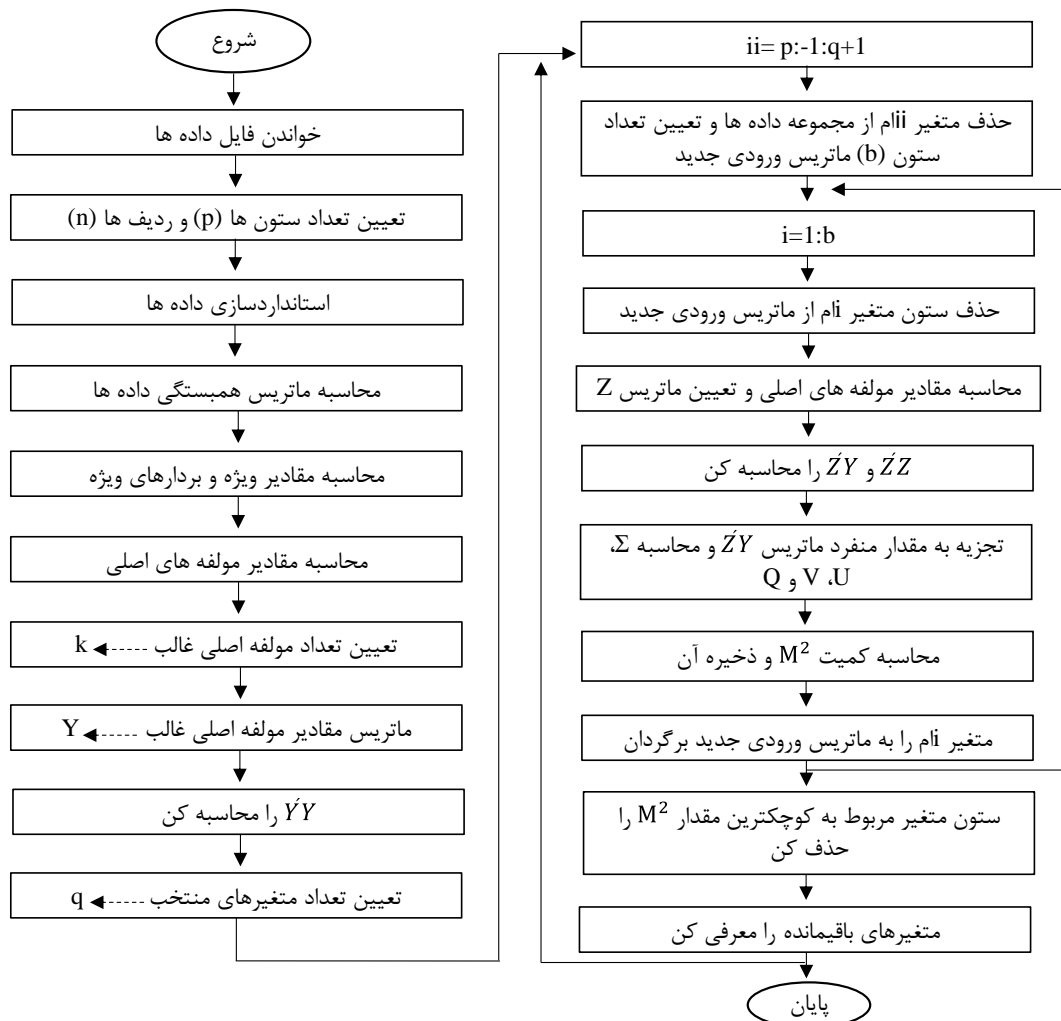
که در آن Q از رابطه ۲ بدست می‌آید، که در آن ماتریس U و V از عملیات تجزیه مقدار منفرد ماتریس $Z' Y$ به ابعاد $k \times k$ حاصل می‌شود که شکل ریاضی آن در فرمول ۳ نشان داده شده است.

$$Q = V U' \quad (2)$$

$$Z' Y = U \Sigma V' \quad (3)$$

که در آن روابط $V V' = I_k$ ، $U U' = I_k$ و $\Sigma = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_k)$ برقرار می‌باشند.

برای هر یک از زیرمجموعه‌های مربوط به مجموعه مرجع متغیرها که حداقل q عضو داشتند، کمیت M^2 تعیین



شکل ۳. مراحل کلی محاسبات روش تجزیه پروکراستس

ماشین بردار پشتیبان یک سیستم یادگیری کارآمد بر مبنای تئوری بهینه‌سازی مقید است که از اصل استقرای کمینه‌سازی خطای ساختاری استفاده کرده و منجر به یک جواب بهینه کلی می‌گردد. در مدل رگرسیون بردار پشتیبان تابعی مرتبط با متغیر وابسته y که خود تابعی از چند متغیر مستقل x است، برآورد می‌شود. مشابه سایر مسائل رگرسیونی فرض می‌شود رابطه میان متغیرهای مستقل و وابسته با تابع جبری مانند $f(x)$ به علاوه مقداری اغتشاش (خطای مجاز ε) مشخص شود.

ماشین بردار پشتیبان یک سیستم یادگیری کارآمد بر مبنای تئوری بهینه‌سازی مقید است که از اصل استقرای کمینه‌سازی خطای ساختاری استفاده کرده و منجر به یک جواب بهینه کلی می‌گردد. در مدل رگرسیون بردار پشتیبان تابعی مرتبط با متغیر وابسته y که خود تابعی از چند متغیر مستقل x است، برآورد می‌شود. مشابه سایر مسائل رگرسیونی فرض می‌شود رابطه میان متغیرهای مستقل و وابسته با تابع جبری مانند $f(x)$ به علاوه مقداری اغتشاش (خطای مجاز ε) مشخص شود.

$$f(x) = W^T \cdot \phi(x) + b \quad (4)$$

$$y = f(x) + noise \quad (5)$$

می‌توان گفت که حدود ۹۰ درصد از مدل‌های شبکه عصبی که در مسائل هیدرولوژی بکار گرفته شده‌اند از نوع پیشرو هستند (Coulbaly et al., 2000) که مدل شبکه عصبی پرسپترون چند لایه با الگوریتم آموزش لوبنبرگ-مارکواردت (Levenberg-Marquardt (LM)) از مهم‌ترین آن‌ها به شمار می‌رود. در شبکه‌های پیشرو، اتصالات میان گره‌هایی وجود دارد که در لایه‌های متفاوت قرار دارند. همچنین ورودی‌ها در لایه ورودی شبکه نمایش داده شده و عمل تحریک از ورودی به سمت خروجی شبکه شروع می‌شود (ASCE, 2000). لازم به ذکر است که در تحقیق حاضر از تعداد نرون‌های ۱ تا ۲۰ برای لایه پنهان شبکه استفاده گردید تا از این طریق بتوان تاثیر ساختار مختلف شبکه بر عملکرد آن در شبیه‌سازی رواناب را مورد ارزیابی قرار داد. معماری معمول این شبکه مشکل از سه بخش لایه ورودی، پنهان و خروجی است. از آنجا که متداول‌ترین نوع توابع محرک تابع تانژانت سیگموئید است، بنابراین در این مطالعه از آن برای نگاشت اطلاعات از لایه ورودی به لایه پنهان و از تابع محرک خطی برای نگاشت اطلاعات از لایه پنهان به لایه خروجی استفاده شد. لازم به ذکر است فرآیند محاسبات شبکه عصبی مصنوعی بر اساس کدنویسی در محیط متلب انجام شده است.

چنانچه W بردار ضرایب، b ثابت مشخصه‌های تابع رگرسیونی و $\phi(x)$ نیز تابع کرنل باشد، آنگاه هدف پیدا کردن فرم تابعی برای $f(x)$ است. این مهم با واسنجی مدل ماشین بردار پشتیبان توسط مجموعه‌ای از نمونه‌ها (مجموعه واسنجی) محقق می‌شود. برای محاسبه W و b لازم است تابع خطا (معادله ۶) در مدل ماشین بردار پشتیبان رگرسیونی با در نظر گرفتن شرایط مندرج (قیود) در معادله ۷ بهینه شود.

$$W^T \cdot \phi(x_i) + b - y_i \leq \varepsilon + \varepsilon_i^* \quad (6)$$

$$\frac{1}{2} W^T w + C \sum_{i=1}^N \varepsilon_i + C \sum_{i=1}^N \varepsilon_i^* y_i - W^T \cdot \phi(x_i) - b \leq \varepsilon + \varepsilon_i$$

$$\varepsilon_i, \varepsilon_i^* \geq 0, i = 1, 2, \dots, N \quad (7)$$

در معادلات بالا C عددی صحیح و مثبت است که عامل تعیین جریمه در هنگام رخ دادن خطای واسنجی مدل است، $\phi(x)$ تابع کرنل، N تعداد نمونه‌ها و دو مشخصه ε_i و ε_i^* متغیرهای کمبود هستند. در نهایت، تابع رگرسیون بردار پشتیبان را می‌توان به فرم زیر باز نویسی کرد:

$$f(x) = \sum_{i=1}^N \bar{\alpha}_i \cdot \phi(x_i)^T \cdot \phi(x) + b \quad (8)$$

در معادله (۸)، $\bar{\alpha}_i$ میانگین ضرایب لاگرانژ است. محاسبه $\phi(x)$ در فضای مشخصه آن ممکن است بسیار پیچیده باشد. برای حل این مشکل روند معمول در مدل رگرسیون بردار پشتیبان انتخاب یک تابع کرنل است.

مراحل انجام پیش‌بینی جریان

در ابتدا براساس نمودارهای ACF مربوط به هر یک از ایستگاه‌های آب‌سنجی صفاخانه و سته، تعداد ۱۵ جریان تاخیر یافته به‌عنوان حداکثر تعداد ورودی مدل‌ها تعیین

هیدرومتری صفاخانه و سته با استفاده از مدل‌های ANN و SVM و به‌ازای دو مجموعه ترکیب مختلف مورد ارزیابی قرار گرفته است که در ذیل به تشریح آن‌ها پرداخته شده است.

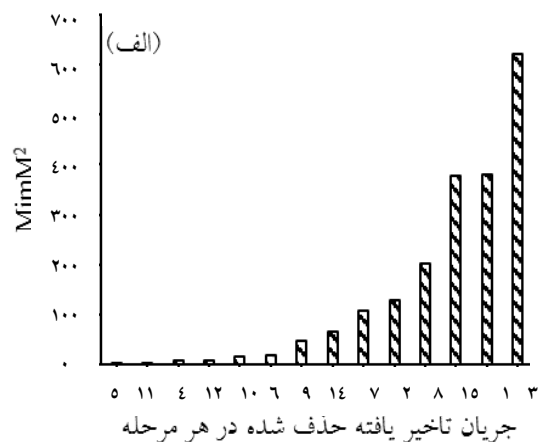
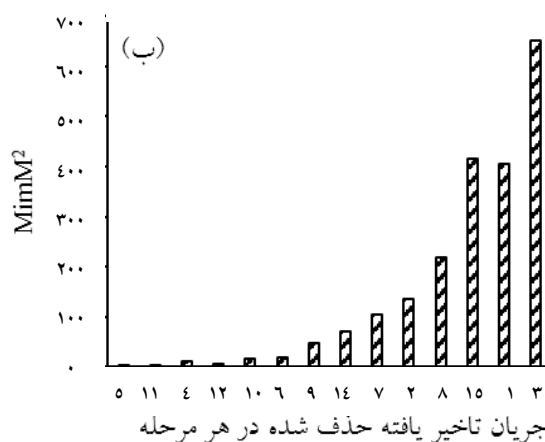
تعیین ورودی با روش تجزیه پروکراستس

براساس داده‌های جریان ماهانه هر یک از دو ایستگاه هیدرومتری صفاخانه و سته ماتریس ورودی روش تجزیه پروکراستس ایجاد گردید. ابعاد این ماتریس $(n \times m)$ می‌باشد که در آن ستون‌ها شامل جریان ماهانه تاخیر یافته $(m=15)$ و ردیف‌ها شامل سری زمانی جریان ماهانه $(n=257)$ می‌باشند. کلیه مراحل نشان داده شده در شکل ۳، به‌ازای k و q برابر ۱ تا ۱۵ انجام گرفته است. روش تجزیه پروکراستس در هر مرحله برابر با عدد تعیین شده برای q ، ورودی‌های موثر را از بین ۱۵ جریان تاخیر یافته تعیین می‌کند. در واقع پس از انجام کلیه مراحل، روش تجزیه پروکراستس ۱۵ ترکیب مختلف برای هر یک از ایستگاه‌های هیدرومتری صفاخانه و سته معرفی می‌کند. در ادامه هر یک از این ترکیب‌ها به‌عنوان ورودی دو مدل شبکه عصبی مصنوعی و ماشین بردار پشتیبان قرار گرفته‌اند. به علت محدودیت صفحات، به‌عنوان نمونه در شکل ۴ مقادیر $MinM^2$ به‌ازای q و k برابر با یک، برای جریان ماهانه ایستگاه‌های هیدرومتری صفاخانه و سته نشان داده شده است.

گردید. سپس مجموعه اول ورودی با ۱۵ ترکیب مختلف، طوری تنظیم گردید که ترکیب اول آن شامل جریان با یک تاخیر، ترکیب دوم شامل جریان با یک و دو تاخیر، ... و ترکیب پانزدهم شامل جریان با یک، دو، ... و پانزده تاخیر باشد. در گام بعدی با استفاده از روش تجزیه پروکراستس مجموعه ترکیب ورودی برای هر یک از ایستگاه‌های آب‌سنجی صفاخانه و سته تعیین گردید. این روش به‌ازای q برابر یک تا پانزده مورد ارزیابی قرار گرفته است، در واقع برای هر یک از ایستگاه‌های آب‌سنجی پانزده ترکیب مختلف معرفی کرده است که این ترکیب‌ها به‌ترتیب شامل یک تا پانزده جریان با تاخیرهای مختلف می‌باشند. سپس مدل‌های شبکه عصبی مصنوعی و ماشین بردار پشتیبان به‌ازای مجموعه اول ورودی و مجموعه ترکیب‌های ارائه شده توسط روش تجزیه پروکراستس، جریان هر یک از ایستگاه‌های آب‌سنجی را مورد پیش‌بینی قرار داده‌اند. دقت و خطای جریان‌های پیش‌بینی شده هر یک از دو مدل شبکه عصبی مصنوعی و ماشین بردار پشتیبان به‌ازای هر یک از ترکیب‌های مورد بررسی، با استفاده از آماره‌های ریشه میانگین مربعات خطا (Root Mean Square Error (RMSE)) و ضریب همبستگی (Correlation Coefficient (CC)) مورد ارزیابی و مقایسه قرار گرفته‌اند.

نتایج و بحث

با توجه به آنچه ذکر شد، پیش‌بینی جریان دو ایستگاه



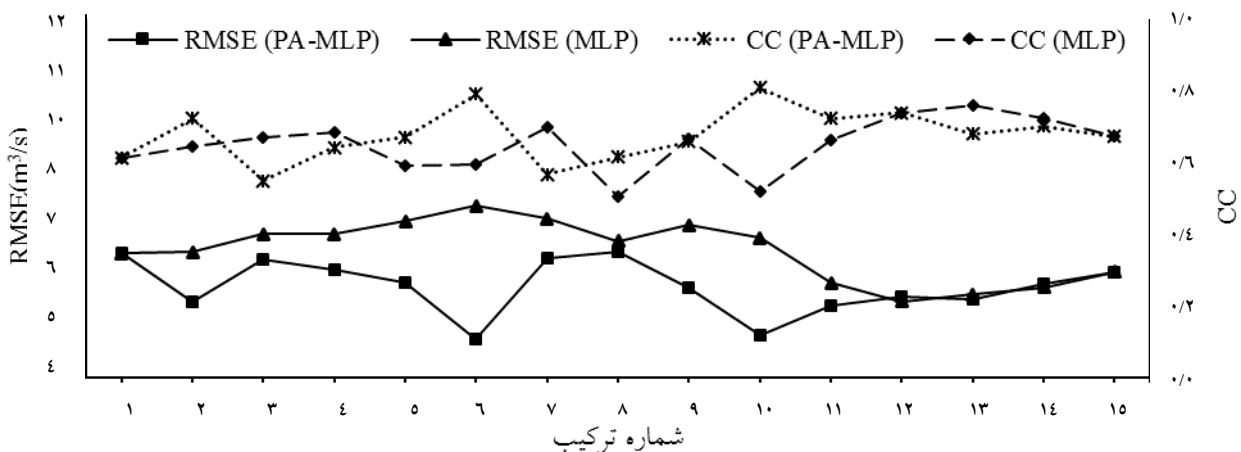
شکل ۴. مقادیر پارامتر $(MinM^2)$ به‌ازای حذف هر یک از جریان‌های تاخیر یافته ایستگاه‌های صفاخانه (الف) و سته (ب)

گردید و سایر ترکیب‌ها نیز به همین ترتیب تعیین گردیده‌اند. در نهایت نیز کلیه ترکیب‌های بدست آمده جهت پیش‌بینی جریان، به عنوان ورودی مدل‌های شبیه‌ساز تعیین گردیده‌اند.

پیش‌بینی جریان ایستگاه صفاخانه نتایج مدل شبکه عصبی مصنوعی

با استفاده از پانزده ترکیب مجموعه اول ورودی و پانزده ترکیب به دست آمده از روش تجزیه پروکراستس، جریان ایستگاه صفاخانه مورد شبیه‌سازی و پیش‌بینی قرار گرفته است. نتایج دوره صحت‌سنجی در شکل ۵ نشان داده شده است.

با توجه به شکل ۴ الف و ب، با حذف ۶ متغیر اول هر یک از ایستگاه‌ها، مجموع اطلاعات موجود در کل داده‌ها تغییر زیادی پیدا نمی‌کند. سپس به‌ازای متغیرهای هفتم تا یازدهم کاهش اطلاعات بیشتر می‌گردد و در متغیرهای دوازده، سیزده و چهاردهم این کاهش اطلاعات به صورت چشمگیری زیاد می‌گردد. به دلیل اینکه در این قسمت هدف تعیین یک متغیر موثر می‌باشد، بنابراین روش پروکراستس برای ایستگاه صفاخانه متغیر اول و برای ایستگاه سنته متغیر سیزدهم را به‌عنوان متغیرهای اصلی تعیین کرده است. به این ترتیب ترکیب اول روش تجزیه پروکراستس برای هر یک از ایستگاه‌های آب‌سنجی تعیین



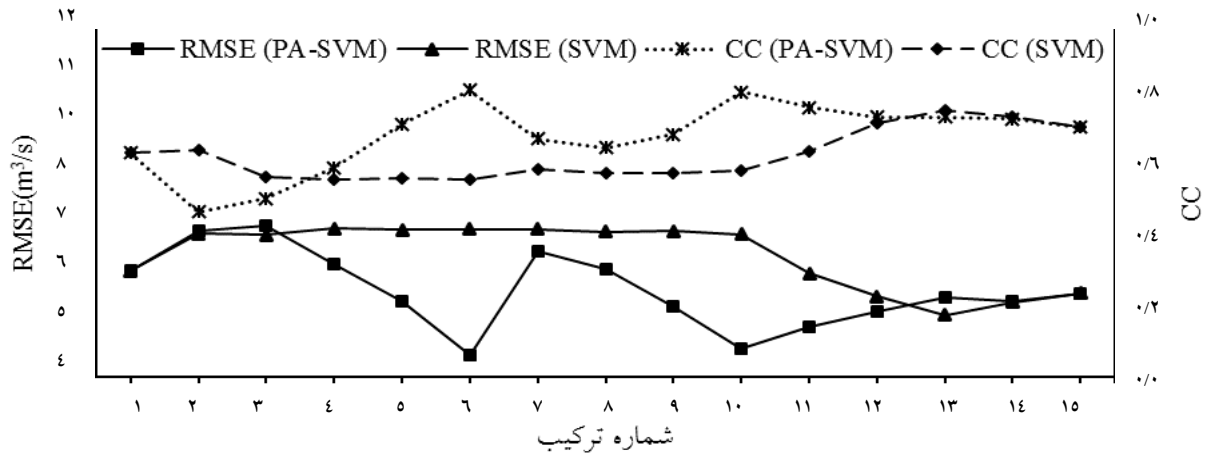
شکل ۵. مقادیر RMSE و CC هر یک از پانزده ترکیب دو مجموعه ترکیب تعیین شده

ماهه جریان رودخانه صفاخانه بوده است. با توجه به نتایج روش تجزیه پروکراستس را می‌توان به‌عنوان یک روش موثر و کارا در تعیین ورودی‌های موثر معرفی نمود.

نتایج مدل ماشین بردار پشتیبان

بهترین نتایج مربوط به هر دو مجموعه ترکیب ذکر شده استخراج گردیده و خلاصه نتایج در شکل ۶ نشان داده شده است.

نتایج نشان می‌دهد که ترکیب شماره دوازده (به‌ازای جریان با تاخیرهای یک تا دوازده ماهه (دوازده ورودی)) با RMSE برابر با $5/68 \text{ (m}^3/\text{s)}$ و CC برابر با $0/73$ بهترین نتیجه را در بین ترکیب‌های ارائه شده مجموعه اول ورودی داشته است. همچنین ترکیب شماره شش ارائه شده توسط روش تجزیه پروکراستس با RMSE و CC به ترتیب برابر با $4/85 \text{ (m}^3/\text{s)}$ و $0/78$ نتیجه بهتری را نسبت به سایر ترکیب‌های ورودی بدست آورده است. ترکیب شماره شش ارائه شده توسط روش تجزیه پروکراستس شامل جریان با تاخیرهای یک، شش، هشت، ده، دوازده و پانزده

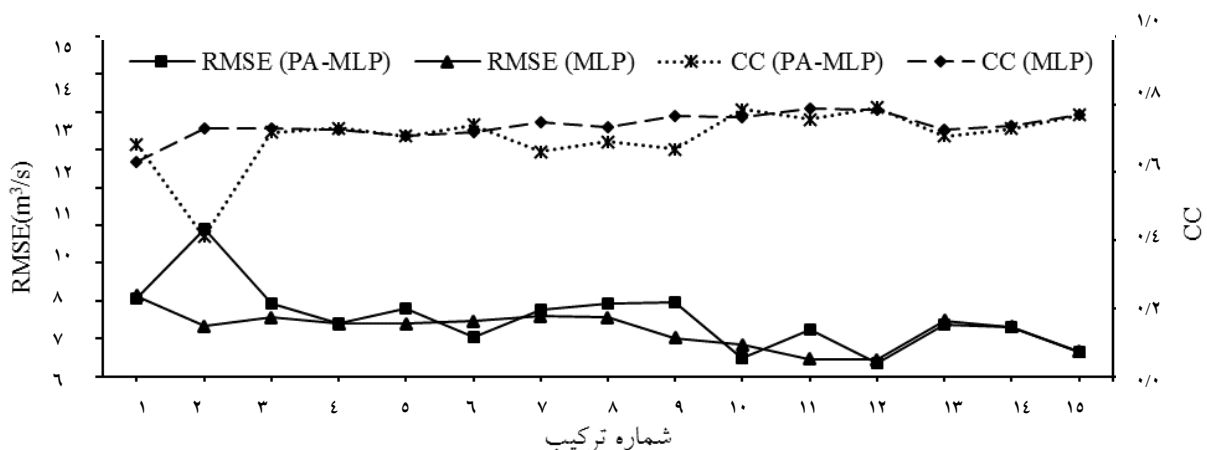


شکل ۶. مقادیر RMSE و CC هر یک از پانزده ترکیب دو مجموعه ترکیب تعیین شده

**پیش‌بینی جریان ایستگاه سنته
نتایج مدل شبکه عصبی مصنوعی**

مقادیر RMSE و CC دوره صحت‌سنجی هر یک از ترکیب‌های ورودی در شکل ۷ نشان داده شده است. نتایج نشان می‌دهد که ترکیب شماره دوازده (به‌ازای دوازده ورودی) با RMSE برابر با $6/44 \text{ (m}^3/\text{s)}$ و CC برابر با $0/78$ بهترین نتیجه را در بین ترکیب‌های ارائه شده در مجموعه اول ورودی داشته است. همچنین ترکیب شماره دوازده ارائه شده توسط روش تجزیه پروکراستس از سایر ترکیب‌های آن با RMSE و CC به ترتیب برابر با $6/36 \text{ (m}^3/\text{s)}$ و $0/79$ بهترین نتیجه را بدست آورده‌اند. ترکیب شماره دوازده ارائه شده توسط روش تجزیه پروکراستس شامل جریان با تاخیرهای یک، دو، چهار، پنج، شش، هفت، نه، ده، یازده، دوازده، چهارده و پانزده می‌باشد.

همانطور که در شکل فوق مشخص است، نتایج اکثر ترکیب‌های بدست آمده از روش تجزیه پروکراستس نسبت به نتایج ترکیب‌های مجموعه اول ورودی بهتر بوده است. ترکیب شماره شش ارائه شده توسط روش تجزیه پروکراستس، دارای نتیجه بهتری نسبت به سایر ترکیب‌های بکار رفته در این بخش می‌باشد. مدل SVM به‌ازای چهار ترکیب دو، سه، سیزده و چهارده نتایج بهتری نسبت به مدل PA-SVM و هر دو مدل به‌ازای ترکیب‌های یک و پانزده (به علت ورودی مشابه) نتایج مشابهی داشته‌اند. پس می‌توان روش تجزیه پروکراستس را یک روش مناسب جهت تعیین ورودی‌های موثر برای مدل SVM معرفی کرد. در پیش‌بینی جریان ایستگاه صفاخانه به‌ترتیب کرنل‌های چندجمله‌ای و پایه‌شعاعی نتایج بهتری را بدست آورده‌اند.

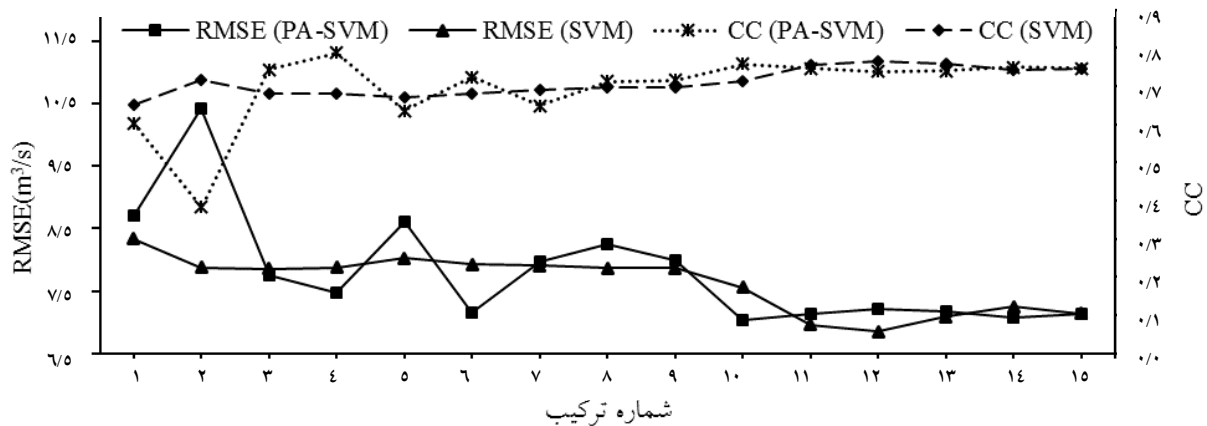


شکل ۷. مقادیر RMSE و CC هر یک از پانزده ترکیب دو مجموعه ترکیب تعیین شده

نتایج مدل ماشین بردار پشتیبان

خلاصه نتایج به‌ازای دو مجموعه ترکیب معرفی شده با استفاده از مدل ماشین بردار پشتیبان در شکل ۸ نشان داده شده است.

لازم به‌ذکر است که در بیشتر ترکیب‌ها، نتایج دو مجموعه ترکیب بسیار به یکدیگر مشابه هستند و نمی‌توان بین دو مجموعه ترکیب ارائه شده توسط روش تجزیه پروکراستس و مجموعه اول ورودی تفاوت معنی‌داری را تشخیص داد.

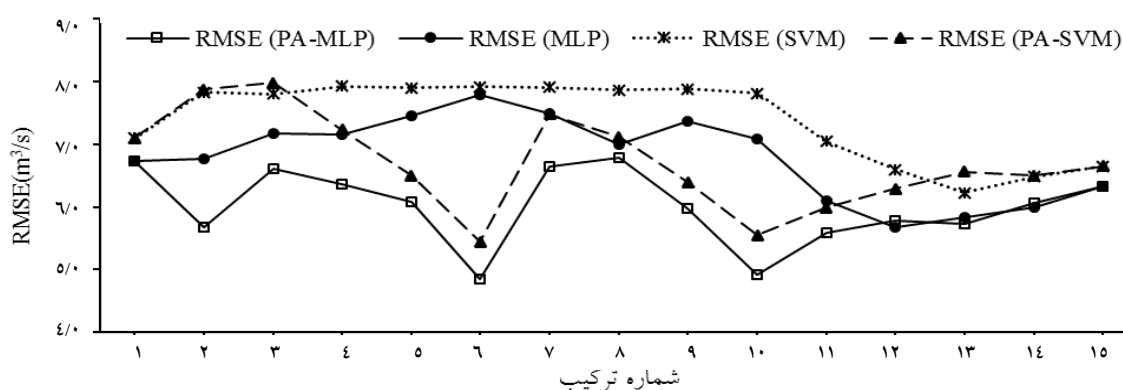


شکل ۸. مقادیر RMSE و CC هر یک از پانزده ترکیب دو مجموعه ترکیب تعیین شده

مقایسه روش‌های ANN و SVM

در گام اول مقایسه براساس نتایج بدست آمده از ترکیب‌های موجود در مجموعه اول ورودی و نتایج ترکیب‌های حاصل از روش تجزیه پروکراستس در پیش‌بینی جریان ایستگاه آب‌سنجی صفاخانه صورت گرفته است. این مقایسه براساس مهم‌ترین معیار مقایسه مدل‌ها، یعنی RMSE دوره صحت‌سنجی انجام گرفته است که خلاصه نتایج در شکل ۹ نشان داده شده است.

نتایج نشان می‌دهد که ترکیب شماره دوازده (به‌ازای دوازده ورودی) با RMSE برابر با $6/85 \text{ (m}^3/\text{s)}$ و CC برابر با $0/77$ بهترین نتیجه را در بین ترکیب‌های ارائه شده مجموعه اول ورودی داشته است. همچنین ترکیب شماره ده ارائه شده توسط روش تجزیه پروکراستس از سایر ترکیب‌های آن با RMSE و CC به ترتیب برابر با $7/03$ و $0/76$ بهترین نتیجه را بدست آورده‌اند. ترکیب شماره ده تعیین شده توسط روش تجزیه پروکراستس شامل جریان با تاخیرهای یک، سه، چهار، شش، هفت، نه، ده، دوازده، سیزده و پانزده می‌باشد. بنابراین در بیش‌تر ترکیبات نتایج هر دو مجموعه ترکیب مشابه یکدیگر می‌باشند و نمی‌توان یکی از مجموعه ترکیب‌ها را به‌عنوان مجموعه ترکیب بهتر معرفی نمود. بدست آوردن ترکیب مناسب بایستی با استفاده از سعی و خطا باشد و روش پروکراستس نیز می‌تواند در برخی مواقع محققان را در بدست آوردن ترکیب مناسب به‌درستی و با دقت بالا راهنمایی کند. همچنین لازم به‌ذکر است که در پیش‌بینی جریان ایستگاه سنته به‌ترتیب کرنل‌های چندجمله‌ای و پایه‌شعاعی نتایج بهتری را بدست آورده‌اند.



شکل ۹. مقایسه عملکرد مدل‌های SVM و ANN در پیش‌بینی جریان ایستگاه صفاخانه به‌ازای دو مجموعه ترکیب ارائه شده

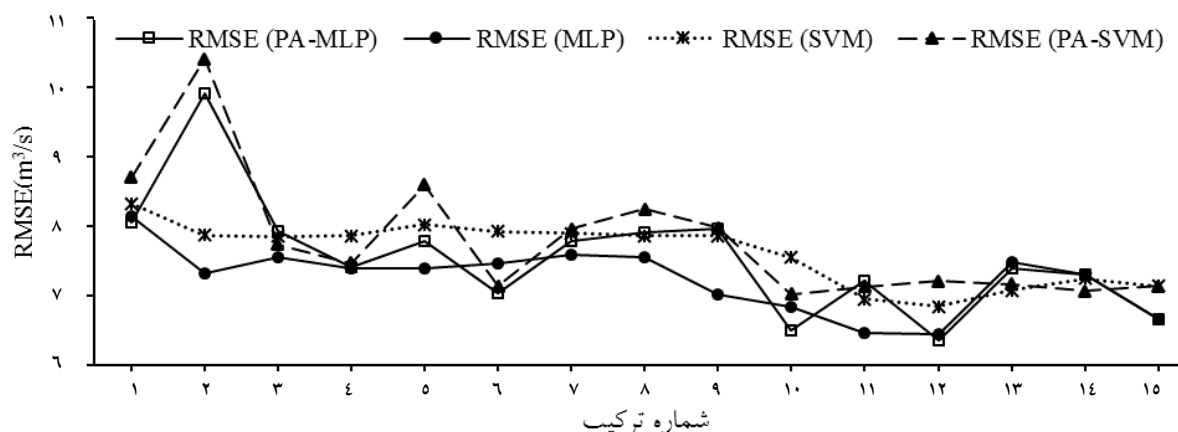
ماشین بردار پشتیبان در پیش‌بینی جریان ایستگاه آب‌سنجی صفاخانه داشته است.

در گام دوم مقایسه براساس نتایج بدست آمده از ترکیب‌های مجموعه اول ورودی و نتایج ترکیب‌های حاصل از روش تجزیه پروکراستس در پیش‌بینی جریان ایستگاه آب‌سنجی سنته صورت گرفته است. مقایسه براساس RMSE دوره صحت‌سنجی انجام گرفته است که در خلاصه آن در شکل ۱۰ نشان داده شده است.

با توجه به شکل فوق مدل MLP به‌ازای کلیه ترکیب‌ها به‌غیر از ترکیب ۱۳ و ۱۴ نسبت به مدل SVM نتایج بهتری را بدست آورده است. همچنین مدل MLP به‌ازای کلیه ترکیب‌ها به‌غیر از ترکیب ۶، ۱۳ و ۱۴ نسبت به مدل (PA-SVM) نتایج بهتری را در پیش‌بینی جریان ایستگاه سنته به خود اختصاص داده است.

با توجه به موارد ذکر شده می‌توان نتیجه گرفت که در هر حالتی می‌توان از مدل شبکه عصبی مصنوعی به‌جای مدل ماشین بردار پشتیبان برای پیش‌بینی جریان ایستگاه آب‌سنجی صفاخانه استفاده کرد.

مدل MLP در کلیه ترکیب‌ها نتایج بهتری را نسبت به مدل SVM و در نه ترکیب نتایج بهتری را نسبت به مدل (PA-SVM) بدست آورده است. البته مدل (PA-MLP) در کلیه ترکیب‌ها نتایج بهتری را نسبت به مدل‌های SVM و (PA-SVM) به دست آورده است. بهترین نتایج مربوط به مدل‌های MLP، SVM، (PA-MLP) و (PA-SVM) به‌ترتیب مربوط به ترکیب‌های دوازده، سیزده، شش و شش می‌باشد که مقادیر RMSE آن‌ها نیز به‌ترتیب برابر با ۵/۶۸، ۶/۲۳، ۴/۸۵ و ۵/۴۵ متر مکعب بر ثانیه می‌باشد. با توجه به آنچه ذکر شده می‌توان نتیجه گرفت که در حالت کلی مدل شبکه عصبی مصنوعی نتایج بهتری را نسبت به مدل



شکل ۱۰. مقایسه عملکرد مدل‌های SVM و ANN در پیش‌بینی جریان ایستگاه سنته به‌ازای دو مجموعه ترکیب ارائه شده

نتایج بهتری را نسبت به مدل SVM و (PA-SVM) بدست آورده است. برتری مدل شبکه عصبی مصنوعی نسبت به مدل ماشین بردار پشتیبان در تحقیقاتی همچون Kisi و همکاران (۲۰۱۲)، Ghorbani و همکاران (۲۰۱۶)، عیسی‌زاده (۱۳۹۴) و در تحقیق حاضر تایید شده است.

روش پروکراستس به علت حذف ورودی‌های مشابه (در نظر گرفتن ورودی‌های کم تشابه‌تر) از مجموعه داده‌های ورودی، توانسته است ورودی‌های موثر را برای پیش‌بینی هر یک از ایستگاه‌های سنجه و صفاخانه پیدا کند. بنابراین در اکثر موارد این روش توانایی تعیین تعداد مشخصی از پارامترهای موثر را دارا است. نتایج این تحقیق نشان داد که نمودار ACF (همبستگی با تاخیرهای متفاوت) به تنهایی نمی‌تواند بطور کامل ورودی‌های موثر را شناسایی کند. ولی روش پروکراستس با حذف ورودی‌های مشابه (با توجه به تعداد ورودی مدنظر) توانسته است دقت مدل‌های ANN و SVM را در تخمین جریان ایستگاه‌ها بیشتر کند. دلیل این امر نیز می‌تواند ناشی از ورودی‌های کم تشابه ورودی استخراج شده با روش پروکراستس می‌باشد. در واقع معرفی چند ورودی مشابه به مدل‌ها سبب افزایش عدم قطعیت مدل‌ها (تنظیم غیر بهینه پارامترهای هر مدل) می‌گردد.

اگرچه در اکثر تحقیقات صورت گرفته در زمینه پیش‌بینی‌های هیدرولوژی عمدتاً از ماشین بردار پشتیبان با تابع پایه شعاعی استفاده می‌شود (اسکندری و همکاران، ۱۳۹۱؛ Khan and Coulibaly, 2006; Yu et al., 2006) اما برتری تابع کرنل چندجمله‌ای در تحقیق حاضر و مطالعاتی از قبیل ازانی و همکاران (۱۳۹۳) و عیسی‌زاده (۱۳۹۴) لزوم بررسی کارایی توابع کرنل مختلف را نشان می‌دهد.

در نهایت نتایج پیش‌بینی‌ها نشان داد که بهترین پیش‌بینی جریان ایستگاه‌های صفاخانه و سنجه مربوط به مدل شبکه عصبی مصنوعی و به‌ازای ورودی‌های بدست آمده از روش پروکراستس می‌باشد. بنابراین روش تجزیه پروکراستس می‌تواند به‌عنوان یک روش مناسب جهت تعیین ورودی‌های موثر در پیش‌بینی جریان خودهمبسته

مدل (PA-MLP) در هشت ترکیب نسبت به مدل SVM و در یازده ترکیب نسبت مدل (PA-SVM) نتایج بهتری را بدست آورده است. بهترین نتایج مربوط به مدل‌های MLP، SVM، (PA-SVM) و (PA-MLP) به ترتیب مربوط به ترکیب‌های دوازده، دوازده، دوازده و ده می‌باشد که مقادیر RMSE آنها نیز به ترتیب برابر با ۶/۴۴، ۶/۸۵، ۶/۳۶ و ۷/۰۳ متر مکعب بر ثانیه می‌باشد. با توجه به آنچه ذکر شده می‌توان نتیجه گرفت که در حالت کلی مدل شبکه عصبی مصنوعی نتایج بهتری را نسبت به مدل ماشین بردار پشتیبان در پیش‌بینی جریان ایستگاه آب‌سنجی سنجه داشته است.

نتیجه‌گیری

پیش‌بینی جریان ایستگاه آب‌سنجی صفاخانه نشان داد که به ترتیب مدل‌های (PA-SVM) و (PA-MLP) در اکثر ترکیب‌ها نتایج بهتری را نسبت به مدل‌های SVM و MLP داشته‌اند.

در پیش‌بینی جریان ایستگاه آب‌سنجی سنجه مدل (PA-MLP) به‌ازای ترکیب شماره دوازده با RMSE و CC به ترتیب برابر با ۶/۳۶ متر مکعب بر ثانیه و ۰/۷۹، بهترین نتیجه را از بین کلیه نتایج بدست آمده با مدل‌های (PA-SVM) و (PA-MLP) داشته است. مدل SVM و (PA-SVM) نیز جریان ایستگاه آب‌سنجی سنجه را با دقت نسبتاً مناسبی پیش‌بینی کرده‌اند. در این بخش از بین کلیه مدل‌های ارزیابی شده، مدل (PA-SVM) به‌ازای ترکیب شماره ده با RMSE و CC به ترتیب برابر با ۷/۰۳ متر مکعب بر ثانیه و ۰/۷۶، بهترین نتیجه را در پیش‌بینی جریان بدست آورده است. مقایسه مدل‌های شبکه عصبی مصنوعی و ماشین بردار پشتیبان به‌ازای ترکیب‌های مجموعه اول ورودی در پیش‌بینی جریان ایستگاه صفاخانه نشان داد که در اکثر مواقع نتایج مدل شبکه عصبی مصنوعی بیشتر از سایر مدل‌ها به واقعیت نزدیک می‌باشد. البته مدل MLP در نه ترکیب نتایج بهتری را نسبت به مدل (PA-SVM) بدست آورده است. همچنین مدل (PA-MLP) در کلیه ترکیب‌ها

معرفی گردد. البته لازم به ذکر است که جهت تعیین بهترین ترکیب ورودی بهتر است از سعی و خطا استفاده گردد و به‌طور قطعی نمی‌توان یک ترکیب ورودی را به‌عنوان بهترین ترکیب ورودی معرفی کرد. همچنین مدل شبکه عصبی مصنوعی می‌تواند به عنوان یک مدل بسیار مناسب جهت پیش‌بینی جریان خودهمبسته مورد استفاده قرار گیرد.

فهرست منابع

- ازانی، ع. ۱۳۹۳. شبیه‌سازی بارش-رواناب با استفاده از ماشین بردار پشتیبان در حوضه آبریز صوفی‌چای. پایان‌نامه کارشناسی ارشد، دانشگاه تبریز، دانشکده کشاورزی، ۱۲۰ صفحه.
- ازانی، ع.، فاضلی‌فرد، م.ح. و قربانی، م.ع. ۱۳۹۳. شبیه‌سازی سطح آب دریاچه ارومیه با استفاده از ماشین‌های بردار پشتیبان و شبکه عصبی مصنوعی. سیزدهمین کنفرانس هیدرولیک ایران، دانشگاه تبریز، ۹۳-۱۰۵.
- اسکندری، ع.، نوری، ر.، معراجی، ع. و کیاقدی، ا. ۱۳۹۱. توسعه مدلی مناسب بر مبنای شبکه عصبی مصنوعی و ماشین بردار پشتیبان برای پیش‌بینی به هنگام اکسیژن خواهی بیو شیمیایی ۵ روزه. مجله محیط‌شناسی، ۳۸(۶۱)، ۷۱-۸۲.
- بنی‌حبیب، م.ا.، ولی‌پور، م. و بهبهانی، س.م.ر. ۱۳۹۰. مقایسه مدل‌های خودهمبسته شبکه عصبی مصنوعی دینامیک و استاتیک در پیش‌بینی جریان ماهانه ورودی به مخزن سد دز. مجله علوم و تکنولوژی محیط زیست، ۱۳(۴)، ۱-۱۴.
- دین‌پژوه، ی. ۱۳۸۲. تحلیل خشکسالی‌های هواشناختی با استفاده از آنالیز الگوها. رساله دکتری، دانشگاه تبریز، دانشکده کشاورزی، ۲۸۸ صفحه.
- عیسی‌زاده، م. ۱۳۹۴. تخمین جریان رودخانه زربینه‌رود با استفاده از مدل‌های هیبریدی فراکاوشی. پایان‌نامه کارشناسی ارشد، دانشگاه تبریز، دانشکده کشاورزی، ۱۷۱ صفحه.
- نوری، ر.، خاکپور، ا.، دهقانی، م. و فرخ‌نیا، ا. ۱۳۸۹. پیش‌بینی ماهانه جریان با استفاده از ماشین بردار پشتیبان بر مبنای آنالیز مؤلفه اصلی. نشریه آب و فاضلاب، ۲۲(۱)، ۱۱۸-۱۲۹.
- Adamowski, J. 2013. Using support vector regression to predict direct runoff, base flow and total flow in a mountainous watershed with limited data in Uttaranchal, India. *Annals of Warsaw University of Life Sciences-SGGW, Land Reclamation*, 45(1): 71-83.
- ASCE Task Committee on Application of Artificial Neural Networks in Hydrology. 2000. Artificial neural networks in hydrology, I: preliminary concepts. *Journal of Hydrologic Engineering*, 5(2): 115-123.
- Asefa, T., Kemblowski, M., McKee, M. and Khalil, A. 2005. Multi-time scale stream flow predictions: The Support vector machines approach. *Journal of Hydrology*, 318(1-4): 7-16.
- Awchi, T.A. 2014. River Discharges Forecasting In Northern Iraq Using Different ANN Techniques. *Water Resources Management*, 28(3): 801-814.
- Coulibaly, P., Anctil, F. and Bobée, B. 2000. Daily reservoir inflow forecasting using artificial neural networks with stopped training approach. *Journal of Hydrology*, 230(3-4): 244-257.
- Dibike, Y., Velickov, S., Solomatine, D. and Abbott, M. 2001. Model induction with of support vector machines: Introduction and applications. *Journal of Computing in Civil Engineering*, 15(3): 208- 216.
- Dinpashoh, Y., Fakhri-Fard, A., Moghaddam, M., Jahanbakhsh, S. and Mirnia, M. 2004. Selection of variables for the purpose of regionalization of Iran's precipitation climate using multivariate methods. *Journal of Hydrology*, 297(1): 109-123.
- Ghorbani, M.A., Ahmadzadeh, H., Isazadeh, M. and Terzi, O. 2016. A comparative study of artificial neural network (MLP, RBF) and support vector machine models for river flow prediction. *Environmental Earth Sciences*, 75(6): 1-14.
- Guo, B., Gunn, S.R., Damper, R.I. and J.D.B. Nelson. 2008. Customizing Kernel Functions for SVM-Based Hyperspectral Image Classification. *IEEE Transactions on Image Processing*, 17(4): 622-629.
- Johnson, G.L. and Hanson, C.L. 1995. Topographic and atmospheric influences on precipitation variability over a mountainous watershed. *Journal of Applied Meteorology*, 34(1): 68-87.
- Jolliffe, I.T. 1986. *Principal Component Analysis*. Springer-Verlag, 271pp.

- Kakaei Lafadani, E., Moghaddam Nia, A., Ahmadi, A., Jajarmizadeh, M. and Ghafari, M. 2013. Stream flow simulation using SVM, ANFIS and NAM models (A Case study). *Caspian Journal of Applied Sciences Reaserch*, 2(4): 86-93.
- Kalteh, A.M. 2013. Monthly river flow forecasting using artificial neural network and support vector regression models coupled with wavelet transform. *Computers and Geosciences*, 54: 1-8.
- Kavzoglu, T. and Colkesen, I. 2009. A kernel functions analysis for support vector machines for land cover classification. *International Journal of Applied Earth Observation and Geoinformation*. 11(5): 352-359.
- Khan, M.S. and Coulibaly, P. 2006. Application of support vector machine in lake water level prediction. *Journal of Hydrologic Engineering*, 11(3): 199-205.
- Kisi, O., Moghaddam Nia, A., Ghafari Gosheh, M., Jamalizadeh Tajabadi, M.R. and Ahmadi, A. 2012. Intermittent Streamflow Forecasting by Using Several Data Driven Techniques. *Water resources management*, 26(2): 457-474.
- Krzanowski, W.J. 1987. Selection of variables to preserve multivariate data structure, using principal components. *Applied Statistics*, 36(1): 22-33.
- Liu, G.Q. 2011. Comparison of Regression and ARIMA models with Neural Network models to forecast the daily stream flow. PhD thesis, University of Delaware. 545pp.
- Misra, D., Oommen, T., Agarwal, A. and Mishra, S.K. 2009. Application and analysis of Support Vector machine based simulation for runoff and sediment yield. *Journal of Biosystems Engineering*, 103(9): 527-535.
- Nam, W., Shin, H., Jung, Y., Joo, K. and Heo, J.H. 2015. Delineation of the climatic rainfall regions of South Korea based on a multivariate analysis and regional rainfall frequency analyses. *International Journal of Climatology*, 35(5): 777-793.
- Nayak, P.C., Sudheer, K.P., Rangan, D.M. and Ramasastri, K.S. 2004. A neuro-fuzzy computing technique for modeling hydrological time series. *Journal of Hydrology*, 291(2-1): 52-66.
- Samsudin, R., Saad, P. and Shabri, A. 2011. River flow time series using least squares support vector machines. *Hydrology and Earth System Sciences*, 15: 1835-1852.
- Terzi, O. and Ergin, G. 2014. Forecasting of monthly river flow with autoregressive modeling and data-driven techniques. *Neural Computing and Applications*, 25(1): 179-188.
- Yu, P.S., Chen, S.T. and Chang, I.F. 2006. Support vector regression for real time flood stage forecasting. *Journal of Hydrology*, 328(3): 704-716.



Autoregressive simulation of Zarrinehrud river basin runoff using Procrustes analysis method and artificial neural network and support vector machine models

Behrooz Sobhani^{1*}, Mohammad Isazadeh² and Monir Shirzad³

^{1*}) Associate Professor, Faculty of Humanities, Climatology, University of Mohaghegh Ardabili, Iran

*Corresponding author email: Sobhani@uma.ac.ir

2) PhD Student, Department of Water Engineering, Faculty of Agriculture, University of Tabriz, Iran

3) MSc Student, Faculty of Geography and Planning, University of Tabriz, Iran

Received: 16-10-2016

Accepted: 24-05-2017

Abstract

Rivers flow prediction in river basins has an important role in the operation and correct management of water resources. Determining type and number of estimator models inputs is one of the important steps in rivers flow prediction. Therefore, The Procrustes analysis (PA) method for determining the number of effective inputs was used. In this study, flow prediction was done using the flow data obtained from the Safakhaneh and Santeh hydrometric stations. The Artificial Neural Network (ANN) and The Support Vector Machine (SVM) models was used for flow prediction. The best estimation of flow is done using the MLP and SVM models in Safakhaneh hydrometric station with RMSE equal to 5.68 (m³/s) and 4.85 (m³/s), respectively, and CC equal to 0.73 and 0.78, respectively. While in Santeh hydrometric station RMSE was equal to 6.44 (m³/s) and 6.36 (m³/s) respectively, and CC was equal to 0.78 and 0.79 respectively for MLP and SVM models. PA-SVM model showed better results than SVM model in estimating Safakhaneh hydrometric stations flow with RMSE equal to 5.45 (m³/s) and CC equal to 0.73 during the test period. The results also indicated that SVM and PA-SVM models estimated the flow of Santeh station with RMSE equal to 6.85 (m³/s) and 7.03 (m³/s) respectively. Basically, results indicated that the Procrustes analysis method can be used as one of the Efficient and suitable methods for determining the number of effective inputs. Comparison of the ANN and SVM results indicated that ANN model has more accuracy than SVM model.

Keywords: Artificial neural network, flow prediction, Procrustes analysis, support vector machine