Archive of SID

نشریه پژوهش‌های حفاظت آب و خاک
جلد بیست و دوم، شماره ششم، ۱۳۹٤
http://jwsc.gau.ac.ir

مقایسه عملکرد روش‌های ماشین بردار پشتیبان و شبکه‌های بیزین در پیش‌بینی جریان روزانه رودخانه (مطالعه موردی: رودخانه باراندوزچای)

***فرشاد احمدی[۱]، فریدون رادمنش[۲] و رسول میرعباسی نجف‌آبادی[۳]**

[۱]دانشجوی دکتری گروه مهندسی منابع آب، دانشگاه شهید چمران، [۲]دانشیار گروه مهندسی منابع آب، دانشگاه شهید چمران اهواز،
[۳]استادیار گروه مهندسی آب، دانشگاه شهرکرد

تاریخ دریافت: ۹۳/۷/۲۹ ؛ تاریخ پذیرش: ۹۴/۳/۱۱

چکیده

سابقه و هدف: از آن‌جا که پیش‌بینی جریان رودخانه‌ها در مهندسی منابع آب از اهمیت ویژه‌ای برخوردار است، پژوهشگران متعددی در دهه‌های اخیر روش‌های مختلفی را در این زمینه به‌کار برده و مورد بررسی قرار داده‌اند. همچنین با توسعه و گسترش روش‌های هوش مصنوعی از یک دهه قبل به این سو، امکان کشف و آشکارسازی بهتر ارتباطات بین متغیرهای مختلف هیدرولوژیکی به‌طور چشم‌گیری میسر گردیده است. ماشین بردار پشتیبان و شبکه‌های بیزین از جمله روش‌های نوینی هستند که در زمینه پیش‌بینی جریان رودخانه در سال‌های اخیر به‌کار گرفته شده‌اند.

مواد و روش‌ها: در این مطالعه، به‌منظور پیش‌بینی جریان رودخانه باراندوزچای از دو روش ماشین بردار پشتیبان (SVM) و شبکه‌های بیزین (BNs) استفاده شد. ماشین بردار پشتیبان یکی از روش‌های یادگیری تحت نظارت است که هم برای دسته‌بندی و هم رگرسیون قابل استفاده است. در سال‌های اخیر سیستم‌های پشتیبانی تصمیم، جهت جمع‌آوری و یکپارچه‌سازی بهترین دانش و اطلاعات برای اتخاذ مناسب‌ترین تصمیم‌گیری، در سطح وسیعی از علوم گسترش یافته است که مدل شبکه تصمیم‌گیری بیزین یکی از انواع مختلف سیستم‌های پشتیبانی تصمیم است. شبکه تصمیم‌گیری بیزین، روشی سودمند است که می‌تواند اطلاعات و داده‌های کمی را به خوبی با دانش کیفی کارشناسی همسو نماید.

یافته‌ها: داده‌های جریان روزانه این رودخانه در محل ایستگاه آبسنجی دیزج در خلال سال‌های ۱۳۸۵ تا ۱۳۸۹ برای ایجاد مدل استفاده شد که ۸۰ درصد داده‌ها برای آموزش و ۲۰ درصد برای صحت‌سنجی مدل به‌کار رفت. همچنین نقش حافظه در کاهش یا افزایش دقت پیش‌بینی مورد بررسی قرار گرفت. جهت مدل‌سازی جریان رودخانه با مدل‌های SVM و BNs از حافظه‌های دبی یک روز قبل، دو روز قبل و ... پنج روز قبل استفاده شد و نتایج براساس شاخص‌های آماری جذر میانگین مربعات خطا، ضریب همبستگی و معیار نش- ساتاکلیف مورد بررسی قرار گرفت. بنابراین الگوهای ورودی متعددی به مدل‌ها معرفی شد. در نهایت مشخص گردید که با افزایش حافظه، دقت مدل‌سازی نیز افزایش می‌یابد به‌طوری‌که بیش‌ترین دقت در الگوی سوم برای هر دو روش به‌دست می‌آید.

* مسئول مکاتبه: farsahad.paper@yahoo.com

www.SID.ir

Archive of SID

مقایسه نتایج دو مدل نشان داد که دقت روش ماشین بردار پشتیبان با R=۰/۹۷۶، RMSE=۱/۸۰ (m^3/s) و معیار نش- ساتاکلیف معادل با ۰/۹۴۹ نسبت به روش شبکه‌های بیزین با R=۰/۹۶۴، RMSE=۱/۹۶ (m^3/s) و معیار نش- ساتاکلیف برابر با ۰/۹۲۱ بیش‌تر است.

نتیجه‌گیری: با دقت بیش‌تر در نتایج ارائه شده، می‌توان نتیجه گرفت که عملکرد دو روش SVM و BNs از نظر تحلیل‌های آماری نزدیک به هم بوده، اما از نقطه نظر گرافیکی و تخمین دبی‌های کمینه و بیشینه، مشخص است که هر دو مدل در پیش‌بینی دبی بیشینه خطای بیش‌تری دارند اما در پیش‌بینی دبی متوسط و کمینه عملکرد مدل SVM به مراتب بهتر از مدل BNs می‌باشد.

واژه‌های کلیدی: شبکه‌های بیزین، ماشین بردار پشتیبان، مدل‌سازی

مقدمه

از آن‌جا که پیش‌بینی جریان رودخانه‌ها در مهندسی منابع آب از اهمیت ویژه‌ای برخوردار است، پژوهشگران متعددی در دهه‌های اخیر روش‌های مختلفی را در این زمینه به‌کار برده و مورد بررسی قرار داده‌اند. روش‌های پیش‌بینی رایج را به‌طور عمده می‌توان به روش‌های مبتنی بر سری‌های زمانی، مدل‌های رگرسیونی و مدل‌های حوضه آبریز تقسیم‌بندی نمود (24). در این بین، روش‌های سری زمانی از محدودیت بیش‌تری برخوردار بوده و در مجموع دقت کم‌تری در پیش‌بینی درازمدت جریان رودخانه دارند. از طرف دیگر، مدل‌های حوضه آبریز که برای پیش‌بینی جریان استفاده می‌شوند بسیار پیچیده بوده و تهیه و تولید آن‌ها نیازمند هزینه بسیار می‌باشد. با توجه به این‌که مدل‌های رگرسیونی می‌توانند ساختارهای ساده و پیچیده‌ای را شامل شوند اما اغلب به‌دلیل پیچیدگی‌های موجود در فرآیند تبدیل بارش و برف به جریان رودخانه قادر به پیش‌بینی دقیق جریان به‌ویژه در فاصله‌های زمانی کوتاه‌مدت نمی‌باشند. از طرفی، کارایی مدل‌های رگرسیونی به‌طور عمده مربوط به پیش‌بینی‌های میان‌مدت بوده و در درازمدت علاوه بر دلایل ذکر شده با توجه به دخالت عوامل دیگر قادر به پیش‌بینی و یا حتی نشان دادن رابطه پارامترهای مختلف و جریان رودخانه نمی‌باشند (23). از سویی دیگر با توسعه و گسترش روش‌های هوش مصنوعی از یک دهه قبل به این سو، امکان کشف و آشکارسازی بهتر ارتباطات بین متغیرهای مختلف هیدرولوژیکی به‌طور چشم‌گیری میسر گردیده است. ماشین بردار پشتیبان[1] و شبکه‌های بیزین[2] ازجمله روش‌های نوینی هستند که در زمینه پیش‌بینی جریان رودخانه در سال‌های اخیر به‌کار گرفته شده‌اند.

ماشین بردار پشتیبان یکی از روش‌های یادگیری تحت نظارت[3] است که هم برای دسته‌بندی و هم رگرسیون قابل استفاده است. این روش توسط وپنیک (1998) بر پایه تئوری یادگیری آماری[4] بنا نهاده شده است (25). ماشین بردار پشتیبان روشی برای طبقه‌بندی دوتایی در فضای ویژگی‌های دلخواه است و از این‌رو روشی مناسب برای مسائل پیش‌بینی به‌شمار می‌رود (18). ماشین بردار پشتیبان در اصل یک دسته‌بندی‌کننده دو کلاسه است که کلاس‌ها را توسط یک مرز خطی از هم جدا می‌کند. در این روش نزدیک‌ترین نمونه‌ها به مرز تصمیم‌گیری را بردارهای

1- Support Vector Machine (SVM)
2- Bayesian Networks (BNs)
3- Supervised Learning
4- Statistical Learning Theory

www.SID.ir

Archive of SID

پشتیبان می‌نامند. این بردارها معادله مرز تصمیم‌گیری را مشخص می‌کنند. الگوریتم‌های شبیه‌سازی هوشمند کلاسیک مانند شبکه‌های عصبی مصنوعی، معمولاً قدرمطلق خطا یا مجموع مربعات خطای داده‌های آموزشی را حداقل می‌کنند، ولی مدل‌های SVM، اصل حداقل‌سازی خطای ساختاری را به‌کار می‌گیرند (9). اخیرا این مدل‌ها در گستره وسیعی از مسایل هیدرولوژیکی و به‌ویژه پیش‌بینی جریان رودخانه‌ها به‌کار رفته‌اند. نیکبخت شهبازی (2008) از مدل ماشین بردار پشتیبان سری‌های زمانی و داده‌های رودخانه‌های کشف رود (ایستگاه آق دربند) 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 می‌کند. مقایسه نتایج حاصل با مدل HEC4، نشان داد که مدل بردار پشتیبان روشی بسیار بهتر و دقیق‌تر در تولید جریان مصنوعی بوده و عملکردی به مراتب بهتر از خود نشان می دهد (17). یو و همکاران (2005) از روش ماشین بردار پشتیبان برای پیش‌بینی سیلاب ساعتی رودخانه لان یانگ در کشور تایوان استفاده کردند. نتایج نشان داد که این روش برای پیش‌بینی سیلاب یک ساعت آینده از دقت بالایی برخوردار بود (26). در مطالعه‌ای دیگر، یو و چن (2006) از روش ماشین بردار پشتیبان جهت پیش‌بینی سیل ساعتی رودخانه لان یانگ در تایوان استفاده نمودند. نتایج حاصل بیانگر دقت مناسب و قابل‌قبول روش مذکور در پیش‌بینی رویداد های سیل در ۱ تا ۶ ساعت بعد داشت (27). آداموفسکی و پراشر (2012) برای پیش‌بینی جریان روزانه در حوضه کوهستانی سیانجی[1] واقع در منطقه هیمالیای هندوستان از دو روش رگرسیون ماشین بردار پشتیبان و شبکه‌های موجک[2] استفاده کردند. مقایسه نتایج نشان داد که گرچه هر دو روش از دقت خوبی در پیش‌بینی جریان روزانه برخوردار بودند، اما دقت روش شبکه‌های موجک اندکی بیش‌تر بود (1). چوبی و همکاران (2014) از مدل SVM برای پیش‌بینی و داده‌کاوری جریان ورودی به سد نارمادا[3] واقع در ایالت مادهایا پرادش[4] هند استفاده کرده و نشان دادند که این روش از توانایی بسیار خوبی در مدل‌سازی و پیش‌بینی جریان میان‌مدت ماهانه برخوردار است (5).

در سال‌های اخیر سیستم‌های پشتیبانی تصمیم[5]، جهت جمع‌آوری و یکپارچه‌سازی بهترین دانش و اطلاعات برای اتخاذ مناسب‌ترین تصمیم‌گیری، در سطح وسیعی از علوم گسترش یافته است که مدل شبکه تصمیم‌گیری بیزین یکی از انواع مختلف سیستم‌های پشتیبانی تصمیم است. شبکه تصمیم‌گیری بیزین[6]، روشی سودمند است که می‌تواند اطلاعات و داده‌های کمی را به خوبی با دانش کیفی کارشناسی همسو نماید (2).

از طرف دیگر این شبکه‌ها را می‌توان به‌عنوان هوش مصنوعی نیز قلمداد کرد. تکنیک‌های مدل‌سازی بیزین دارای ویژگی‌هایی هستند که آن‌ها را برای پرسش‌های تحلیلی و مدیریتی داده‌های جهان واقعی، کاربردی و مفید ساخته‌اند (20). امروزه این مدل‌های احتمالاتی کاربرد وسیعی در علوم مربوط به محیط زیست و منابع آب پیدا کرده‌اند. برای مدل بیزین می‌توان مزیت‌های زیر را بیان کرد:

۱) در شبکه‌های بیزین جهت انجام تحلیل، چیزی به‌عنوان اندازه حداقل برای داده‌های نمونه وجود

1- Sianji

2- Wavelet Networks
3- Narmada
4- Madhya Pradesh
5- Decision Support System (DSS)
6- Bayesian Decision Network (BDN)

www.SID.ir

Archive of SID

ندارد و این شبکه‌ها در محاسبات خود از تمامی داده‌های موجود استفاده می‌کنند و حتی با وجود نمونه‌هایی با اندازه‌های نسبتاً کوچک می‌توان به پیش‌بینی‌هایی با دقت خوب دست یافت. ۲) به‌کارگیری و ترکیب منابع مختلف داده‌ها: از مهم‌ترین ویژگی‌های شبکه بیزین، استفاده از اطلاعات پیشین و اولیه راجع به موضوع مورد نظر می‌باشد. سپس جهت کسب ترکیبی از دانش اولیه و داده‌های جدید، این اطلاعات به‌وسیله داده‌ها به روز می‌شوند. این تلفیق به‌دست آمده نیز می‌تواند به‌عنوان اطلاعات اولیه در یک مطالعه جدید مورد استفاده قرار گیرد. از دیگر امتیازات شبکه‌های بیزین این است که به راحتی در یک مسیر ریاضی، می‌تواند اطلاعات به‌دست آمده با درجه دقت و منابع مختلف را با هم ترکیب کرده و مورد استفاده قرار دهد. در این صورت دانش کارشناسی می‌تواند در خصوص متغیرهایی که داده لازم در مورد آن‌ها موجود نیست، با سایر اطلاعات در دسترس ترکیب شود (15). ۳) توان بالای شبکه‌های بیزین در نمایش تصویری متغیرها و روابط بین آن‌ها در یک سیستم منجر به شناخت و درک بالاتری از سامانه می‌گردد (21).

رجینی و ویترز (2008) با کاربرد مدل‌های بیزین و با در نظر گرفتن عدم قطعیت‌های موجود به پیش‌بینی تراز سطح آب در سیستم پیش‌بینی جریان رود راین پرداختند. با ترکیب یک تابع هزینه در مدل برای برآورد ارزش اقتصادی ارسال هشدار سیلاب، خسارت ناشی از هر تراز سطح آب پیش‌بینی شده را تعیین کرده و به این نتیجه رسیدند که به کمک این روش، با یک ریسک قابل‌قبول می‌توان تصمیم‌گیری در شرایط عدم قطعیت را برای پیش‌بینی جریان رودخانه تسهیل نمود (20). مددگر و مرادخانی (2014) در مطالعه‌ای به پیش‌بینی تغییرات مکانی خشکسالی در حوضه آبریز رودخانه گانیسون در آمریکا پرداختند. بدین منظور برای پیش‌بینی رواناب حوضه و محاسبه خشکسالی‌های آتی، از مدل‌های بیزین استفاده شده بود. نتایج حاصل نشان داد که این مدل‌ها در پیش‌بینی جریان رودخانه از توانایی بالایی برخوردار هستند (13).

با عنایت به موارد فوق می‌توان دریافت که پیش‌بینی متغیرهای هیدرولوژیکی هم‌چون جریان رودخانه‌ها از دیرباز مورد توجه پژوهشگران امر بوده و بدین منظور روش‌های متعددی از جمله مدل‌های تجربی- نیمه‌تجربی، سری‌های زمانی و مدل‌های هوشمند توسعه یافته‌اند که در این میان مدل‌های هوشمند با الهام گرفتن از طبیعت، از عملکرد بهتری نسبت به سایر روش‌ها برخوردار هستند. همچنین با توجه به بررسی منابع صورت گرفته مشاهده می‌شود که در منابع داخلی از مدل‌های شبکه بیزین و ماشین بردار پشتیبان به‌صورت بسیار محدودی در پیش‌بینی متغیرهای هیدرولوژیکی استفاده شده است. بنابراین در این پژوهش سعی گردیده است که عملکرد دو مدل هوشمند BNs و SVM در پیش‌بینی جریان روزانه رودخانه باراندوزچای مورد بررسی و مقایسه قرار گیرد.

مواد و روش‌ها

معرفی منطقه مورد مطالعه: در این مطالعه، از داده‌های 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 شمال‌غرب کشور بین دریاچه ارومیه و مرز ایران و کشورهای عراق و ترکیه در موقعیت جغرافیایی ´۴۵ °۴۴ تا ´۱۴ °۴۵ طول شرقی و ´۰۶ °۳۷ تا ´۲۹ °۳۷ عرض شمالی واقع شده است. طول آبراهه اصلی حوضه ۷۵ کیلومتر بوده و

www.SID.ir

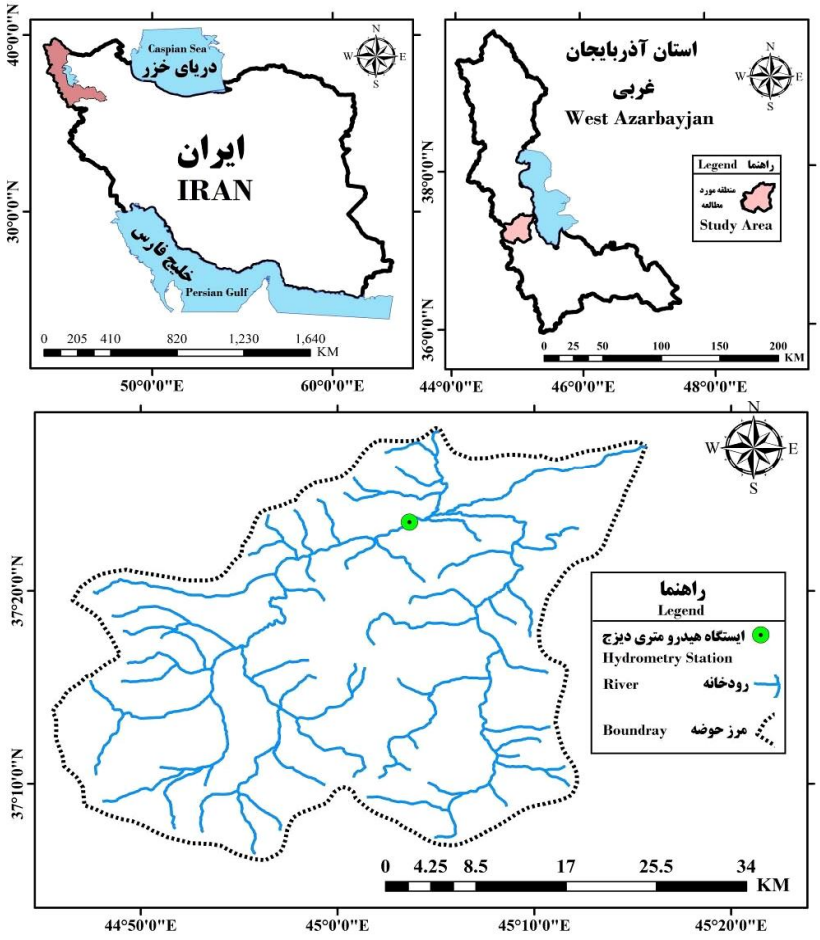
Archive of SID

حداکثر و حداقل ارتفاع این حوضه به‌ترتیب ۳۵۰۰ و ۱۲۵۰ متر از سطح آزاد دریاها می‌باشد. در جدول ۱ مشخصات آماری ایستگاه هیدرومتری دیزج آورده شده است. شکل ۱ نیز حوضه آبریز باراندوزچای و موقعیت ایستگاه هیدرومتری دیزج را نشان می‌دهد.

جدول ۱- مشخصات آماری سری جریان روزانه رودخانه باراندوزچای در محل ایستگاه دیزج.

Table 1. Basic statistics of the Baranduz Chai daily river flow in Dizaj Station.

مشخصات آماری دبی روزانه (Basic Statistics of daily river flow)	(m^3/s)
میانگین (Mean)	5.92
واریانس (Variance)	57.98
انحراف معیار (Standard deviation)	7.61
حداکثر (Max.)	92.60
حداقل (Min.)	0.00



شکل ۱- موقعیت حوضه آبریز باراندوزچای و ایستگاه هیدرومتری دیزج.

Figure 1. Location of Baranduz Chai river basin and Dizaj hydrometric station.

www.SID.ir

Archive of SID

شبکه‌های بیزین: شبکه بیزین یک مدل گرافیکی احتمالاتی است که مجموعه‌ای از متغیرها و احتمالات مربوط به هر کدام را نشان می‌دهد. این شبکه یک گراف مستقیم و غیرچرخه‌ای است که در آن گره‌ها در حکم متغیرهای مسأله هستند. شبکه‌های بیزین محدودیتی در نمایاندن متغیرهای تصادفی ندارند. ساختار یک شبکه بیزین در واقع یک نمایش گرافیکی از اثرات متقابل متغیرهایی است که باید مدل شوند و علاوه بر این که کیفیت رابطه بین متغیرهای مسأله را نشان می‌دهد، کمیت ارتباط بین این متغیرها را نیز به نمایش می‌گذارد که به‌صورت عددی از توزیع احتمال مشترک آن‌ها استفاده می‌کند (7). این روش بر مبنای محاسبات احتمالات وابسته یا قانون بیز می‌باشد (رابطه ۱). یک شبکه بیزی را می‌توان این‌گونه تعریف کرد: تعدادی گره که نشان‌دهنده آن دسته از متغیرهای تصادفی هستند که با یکدیگر در تعامل می‌باشند. این برهم‌کنش، به‌وسیله ایجاد ارتباط بین گره‌ها ایجاد می‌شود (3).

$$P(b|a) = P(a|b) \times \frac{P(b)}{P(a)} \qquad (۱)$$

که در آن، $P(a)$ احتمال وقوع پیشامد a، $P(b)$ احتمال وقوع پیشامد b، $P(b|a)$ احتمال وقوع پیشامد b به شرطی که a اتفاق افتاده باشد و $P(a|b)$ احتمال وقوع پیشامد a به شرطی که b اتفاق افتاده باشد، هستند.

متغیرهایی که در مسایل زیست‌محیطی و منابع آب به‌کار می‌روند، می‌توانند کیفی یا کمی باشند. تعیین تعداد مناسبی از متغیرها در یک مسأله مدل‌سازی از مسایل چالش‌برانگیزی است که باید با دقت به آن پرداخته شود. قانون احتمالات اعمال شده نیز باید بر مبنای بهترین یافته‌های موجود یا اطلاعات کمی و کیفی بین متغیرها باشد. پیشنهادها و تصمیم‌هایی که برای اجرا ارائه می‌شوند، بر اساس حداکثر سود پیش‌بینی شده است (15). هر شبکه بیزین از سه جزء اصلی تشکیل شده است:

۱) مجموعه‌ای از گره‌ها که در واقع متغیرهای سیستم مدیریتی هستند. این گره‌ها می‌توانند متغیرهای پیوسته یا ناپیوسته، اعداد ثابت و یا توابع پیوسته باشند. هر گره به چند طبقه یا کلاس تقسیم می‌شود. در حالت کلی گره‌ها یا والد هستند یا ولد (فرزند). یک گره ولد می‌تواند از چندین والد تولید شده باشد (14)

۲) مجموعه‌ای از بندها: روابط سببی بین متغیرها به‌وسیله بندها به وجود می‌آیند (21) این روابط به‌وسیله پیکان نمایش داده می‌شوند. گره‌ها به‌وسیله پیکان‌ها به هم مربوط می‌شوند. عدم وجود پیکان بین گره‌ها، بیانگر استقلال متغیرها می‌باشد (19). گره‌هایی که هیچ پیکانی به آن‌ها وارد نشده است، گره‌های والد ورودی هستند. گرهی که پیکان به آن وارد و از آن خارج می‌شود، بیانگر گره ولد وضعیت می‌باشند و گره‌هایی که از آن‌ها هیچ پیکانی خارج نمی‌شود، نشان‌دهنده گره‌های ولد خروجی می‌باشند. منظور منطقی پیکانی که از سمت متغیر x به‌سمت متغیر y میرود، این است که متغیر x اثر مستقیمی بر متغیر y دارد. مثلاً دما اثر مستقیم بر تبخیر و تعرق می‌گذارد (19).

۳) مجموعه‌ای از احتمالات: هر کدام برای متغیر شرایط تصمیمی را مشخص می‌کند که این شرایط از طریق متغیرهایی که مستقیماً آن را تحت‌تأثیر قرار می‌دهند (والدین آن)، به آن نسبت داده می‌شوند (3). گره‌هایی که قبل از آن‌ها گره دیگری در گراف وجود دارد با توزیع احتمالی شرطی تعریف می‌شوند و در

www.SID.ir

Archive of SID

غیر این صورت، با احتمال آغازین (اولیه) بیان می‌شوند. احتمال شرطی، احتمال میزان تأثیر یک گره ورودی به یک گره مفروض در شبکه را به‌دست می‌دهد و احتمال آغازین بیانگر میزان احتمال قرارگیری یک متغیر ورودی در یک کلاس معین می‌باشد و از طریق اطلاعات اولیه موجود راجع به متغیرها به‌دست می‌آید. احتمالات مربوط به پایین‌ترین بخش در شبکه بیزین از طریق قانون احتمال کل به‌دست می‌آید و احتمالات مرتبط با بخش‌های بالایی شبکه، براساس قانون بیز هستند. در حالت کلی، اگر گره x دارای والد نباشد گره غیرمشروط و در غیر این صورت گره مشروط می‌باشد (12).

ماشین‌های بردار پشتیبان: در یک مدل رگرسیونی SVM لازم است وابستگی تابعی متغیر وابسته y به مجموعه‌ای از متغیرهای مستقل x تخمین زده شود. فرض بر این است که مانند دیگر مسایل رگرسیونی، رابطه بین متغیرهای وابسته و مستقل توسط یک تابع معین f به‌علاوه یک مقدار اضافی نویز (Noise) مشخص می‌شود.

$$y = f(x) + Noise \qquad (۲)$$

بنابراین موضوع اصلی، پیدا کردن فرم تابع f است که بتواند به صورت صحیح، موارد جدیدی را که SVM تاکنون تجربه نکرده است پیش‌بینی کند. این تابع به‌وسیله آموزش مدل SVM بر روی یک مجموعه داده به‌عنوان مجموعه آموزش که شامل فرآیندی به‌منظور بهینه‌سازی دایمی تابع خطا است، قابل دسترسی است. بر مبنای تعریف این تابع خطا، دو نمونه از مدل‌های SVM شناخته شده است که عبارتند از الف) مدل‌های رگرسیونی SVM نوع اول که مدل‌های $SVM - v$ نیز نامیده می‌شوند و ب) مدل‌های رگرسیونی SVM نوع دوم که با نام $SVM - \varepsilon$ شناخته شده هستند. در این مطالعه $SVM - \varepsilon$ به‌دلیل کاربرد گسترده آن در مسایل رگرسیونی مورد استفاده قرار گرفت. برای این مدل، تابع خطا به‌صورت زیر تعریف می‌شود:

$$\frac{1}{2}W^T W + C\sum_{i=1}^{N}\xi_i + C\sum_{i=1}^{N}\xi_i^* \qquad (۳)$$

تابع خطای فوق لازم است که با توجه به محدودیت‌های زیر کمینه گردد (9):

$$\begin{aligned} & W^T\phi(x_i) + b - y_i \le \varepsilon + \xi_i^* \\ & y_i - W^T\phi(x_i) - b \le \varepsilon + \xi_i^* \\ & \xi_i, \xi_i^* \ge 0 \end{aligned} \qquad (۴)$$

که در آن، 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 می‌باشد. ماشین‌های بردار پشتیبان برای حل مسایل غیرخطی، ابعاد مسأله را از طریق توابع کرنل تغییر می‌دهند. انتخاب کرنل برای SVM به حجم داده‌های آموزشی و ابعاد بردار ویژگی، بستگی دارد. به‌عبارت دیگر، باید با توجه به این پارامترها تابع کرنلی را انتخاب نمود که توانایی آموزش برای ورودی‌های مساله را داشته باشد. در عمل چهار نوع کرنل خطی[1]، کرنل چندجمله‌ای[2]، کرنل تانژانت هیپربولیک[3] و کرنل گوسی[4] به‌کار گرفته می‌شوند. در جدول ۲ معادلات برخی از کرنل‌های رایج ارائه شده‌اند.

1- Linear kernel
2- Polynomial kernel
3- Hyperbolic Tangent Kernel
4- Radial Base Function kernel

www.SID.ir

Archive of SID

جدول ۲- توابع کرنل رایج در ماشین‌های بردار پشتیبان (9).

Table 2. Common kernel function in SVMs (9).

نوع تابع (Function)	تابع (Kernel Function)
خطی (Linear)	$K(x_i,x_j)=x_i^T.x_j$
چندجمله‌ای (Polynomial)	$K(x_i,x_j)=(\gamma\ x_i^T.x_j+C)^d$
تانژانت هیپربولیک (Hyperbolic Tangent)	$K(x_i,x_j)=\tanh(\gamma\ x_i^T.x_j+C)$
RBF	$K(x_i,x_j)=\exp\left(-\gamma\left\|x_i-x_j\right\|^2\right)$

در نهایت، تابع تصمیم رگرسیون بردار پشتیبان غیرخطی به‌صورت معادله زیر خواهد بود که کنترل‌کننده میزان نوسان تابع گوسی و همچنین کنترل‌کننده نتایج پیش‌بینی و تعمیم‌دهنده مدل SVM است (چن و یو، ۲۰۰۶).

$$f(x_i)=\sum_{i=l}^{l}(-\partial_i-\partial_i^*)K(x_i,x_j)+b \qquad (۵)$$

معیارهای ارزیابی مدل: در این پژوهش، برای ارزیابی مدل‌های مورد نظر از معیارهای ضریب همبستگی، جذر میانگین مربعات خطا و معیار نش- سالتاکلیف (E) استفاده شد:

$$R=\frac{\sum_{i=1}^{n}(Q_i-\overline{Q_i})(Q_{For}-\overline{Q_{For}})}{\sqrt{\sum_{i=1}^{n}(Q_i-\overline{Q_i})^2\sum_{i=1}^{n}(Q_{For}-\overline{Q_{For}})^2}} \qquad (۶)$$

$$RMSE=\left(\frac{\sum_{i=1}^{n}(Q_i-\overline{Q_{For}})^2}{n}\right)^{0.5} \qquad (۷)$$

$$E=\left(1-\frac{\sum_{i=1}^{n}(Q_i-\overline{Q_{For}})^2}{\sum_{i=1}^{n}(Q_i-\overline{Q_i})^2}\right) \qquad (۶)$$

که در آن، Q_i مقدار مشاهده شده، Q_{For} مقدار محاسبه شده، n تعداد داده‌ها، $\overline{Q_i}$ میانگین مقادیر مشاهداتی و $\overline{Q_{For}}$ میانگین مقادیر محاسباتی می‌باشد. هر چه مقدار ضریب همبستگی بالاتر و RMSE کم‌تر باشد، نتایج دقیق‌تر و قابل اعتمادتر خواهند بود. معیار نش- ساتکلیف برای بررسی بازده مدل کاربرد داشته و می‌تواند مقداری بین یک تا ∞ را اختیار کند. هرچه مقدار این معیار به عدد یک نزدیک‌تر باشد، نشان‌دهنده کارایی بیش‌تر مدل است (22).

نتایج و بحث

نتایج مدل‌سازی جریان روزانه رودخانه باراندوزچای با استفاده از شبکه بیزین: در مطالعه حاضر، برای مدل‌سازی جریان روزانه رودخانه باراندوزچای با استفاده از مدل BNs، 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) به‌عنوان داده‌های صحت‌سنجی انتخاب شدند (به‌طورکلی ۸۰ درصد داده‌ها برای آموزش و ۲۰ درصد برای صحت‌سنجی در نظر گرفته شد). در مدل‌های هوشمند انتخاب ورودی‌های اولیه

www.SID.ir

Archive of SID

مناسب و تأثیرگذار در پدیده به‌منظور آموزش ماهیت ساز و کار حاکم بر پدیده باعث بهبود عملکرد خواهد شد. بنابراین در الگوسازی جریان رودخانه باراندوزچای نیز سعی گردید تا موثرترین داده‌های مشاهداتی به‌عنوان داده‌های آموزشی انتخاب شود. در جدول ۳ الگوهای مختلف ورودی مدل BNs ارائه شده است. با توجه به این جدول از نقش حافظه (ارتباط دبی امروز با جریان روزهای ماقبل) در مدل‌سازی جریان استفاده شده و از اضافه کردن سایر متغیرهای هیدرولوژیکی که نقش کمی در جریان رودخانه داشته و صرفاً هزینه جمع‌آوری داده و زمان صرف شده را افزایش می‌دهد پرهیز گردید. این الگوها شامل دبی یک روز قبل ((Q(t-1)) تا دبی پنج روز قبل ((Q(t-5)) می‌باشند. البته در مطالعات مختلفی هم‌چون فربودنام و همکاران (2008) و داننده‌مهر و مجدزاده طباطبایی (2010) نیز به مانند مطالعه حاضر از حافظه برای پیش‌بینی جریان رودخانه استفاده شده است (به‌دلیل کاهش زمان، هزینه و پیچیدگی مدل‌ها) (8, 6).

مراحل ساخت و توسعه مدل شبکه بیزین در مطالعه حاضر برای مدل‌سازی جریان روزانه رودخانه باراندوزچای به‌صورت زیر انجام شد:

۱- تشکیل چارچوب مدل: که شامل تعریف متغیرها و روابط موجود بین آن‌ها در سیستم است و معمولاً براساس یک مدل مفهومی از سیستمی که قرار است مدل‌سازی و مدیریت شود، شکل می‌گیرد. بنابراین ساخت مدل مفهومی، اولین قدم در ساختن مدل بیزین است. در این مطالعه، مدل مفهومی بیزین از طریق الگوهای ورودی ساخته شد (شکل ۲). ۲) تعیین نوع ارتباط بین ورودی‌های مختلف مدل در ساختار مفهومی: در مطالعه حاضر از حافظه سری روزانه جریان به‌عنوان ورودی‌های اصلی استفاده شد. حال در تعریف مفهومی برای مدل بیزین این گونه در نظر گرفته شد که، دبی فردا متأثر از توالی‌های گذشته خود می‌باشد (شکل ۲).

۳) آموزش شبکه و به‌دست آوردن رابطه بین داده‌ها: پس از تشکیل چارچوب و تعریف مدل، نوبت به آموزش شبکه ساخته شده می‌رسد. در این مرحله با استفاده از الگوهای مختلف ورودی (الگوهای ارائه شده در جدول ۳)، مدل سعی در ایجاد ارتباط بین داده‌ها داشته و شکل ریاضی این ارتباط را به‌دست می‌آورد. حال با رابطه ریاضی موجود می‌توان داده مصنوعی تولید کرده و عملکرد مدل را مورد ارزیابی قرار داد.

۴) ارزیابی مدل: مجوعه‌ای از ابزار اعتبارسنجی می‌تواند در مورد ارزیابی شبکه به‌کار برده شود که این ارزیابی‌ها می‌توانند از طریق داده‌ها و معیارهای ارزیابی موجود در مواد و روش‌ها صورت گیرد.

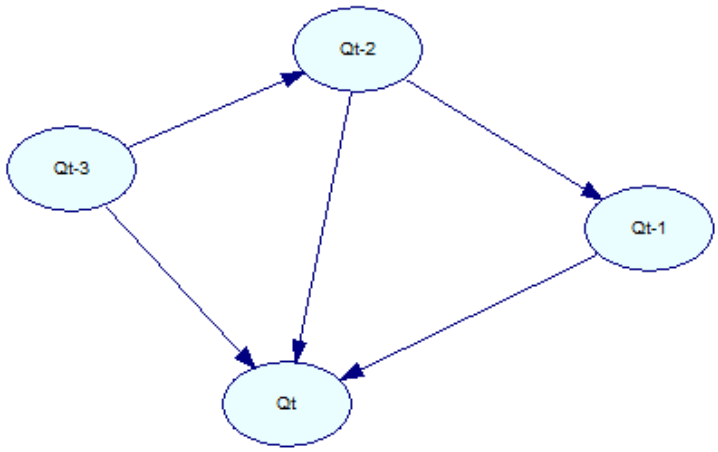
جدول ۳- الگوهای ورودی مورد استفاده در مقیاس زمانی روزانه رودخانه باراندوزچای.

Table 3. Input patterns used in the daily time scale of Baranduz Chai River.

ردیف (Row)	مدل (Model)	الگوی ورودی جریان روزانه (Daily iputs)
۱	M_1	Q(t)=f{Q(t-1)}
۲	M_2	Q(t)=f{Q(t-1), Q(t-2)}
۳	M_3	Q(t)=f{Q(t-1), Q(t-2), Q(t-3)}
۴	M_4	Q(t)=f{Q(t-1), Q(t-2), Q(t-3), Q(t-4)}
۵	M_5	Q(t)=f{Q(t-1), Q(t-2), Q(t-3), Q(t-4), Q(t-5)}

www.SID.ir

Archive of SID



شکل ۲- مدل مفهومی ساخته شده برای شبکه بیزین.

Figure 2. Bayesian networks conceptual model.

جدول ۴ نتایج ارزیابی الگوهای مختلف ورودی شبکه بیزین برای مدل‌سازی جریان روزانه رودخانه در مراحل آموزش و صحت‌سنجی را نشان می‌دهد. با توجه به این جدول می‌توان نتیجه گرفت که مدل شامل دبی سه روز قبل (M_3) با R=۰/۹۶۴، RMSE=۱/۹۶(m^3/s) و معیار نش- ساتاکلیف معادل با ۰/۹۲۱ دارای بیش‌ترین دقت در مرحله صحت‌سنجی بوده و به‌عنوان بهترین الگو برای مدل بیزن انتخاب می‌شود. همچنین رابطه ریاضی حاصل از الگوی سوم به‌صورت رابطه (۸) به‌دست آمد.

$$Q_t=0.265818Q_{t-3}-0.0417711Q_{t-2}+0.717615Q_{t-1} \quad (۸)$$

شکل ۳ نمودار مقادیر مشاهداتی و مقادیر پیش‌بینی شده الگوی بهینه (M_3) برای سری روزانه جریان رودخانه باراندوزچای، در مرحله صحت‌سنجی را نشان می‌دهد.

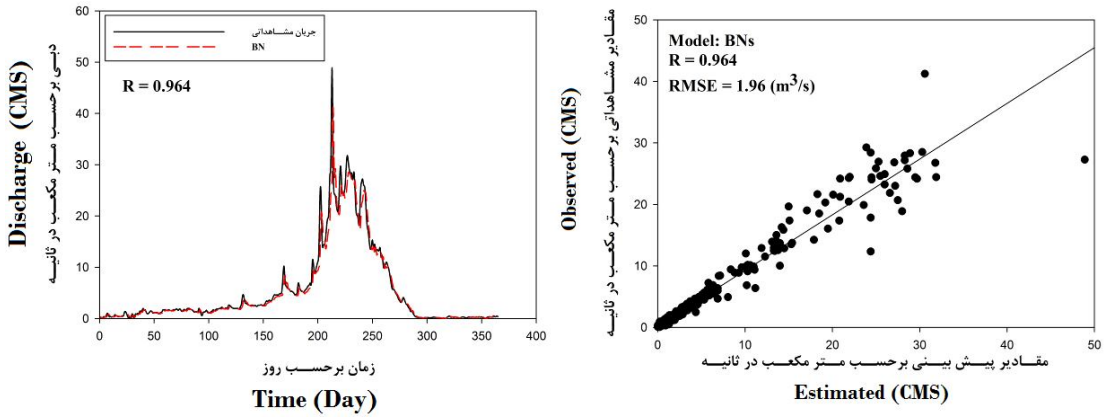
جدول ٤- تحلیل‌های آماری نتایج مدل بیزین برای الگوهای ورودی جریان روزانه.

Table 4. The Statistical analysis of Bayesian model results for daily river flow inputs.

صحت‌سنجی (Test)			آموزش (Train)			الگو (Model)	ردیف (Row)
E	RMSE (m^3/s)	R	E	RMSE (m^3/s)	R		
0.906	2.12	0.957	0.843	3.23	0.908	M_1	1
0.910	2.07	0.959	0.851	3.17	0.910	M_2	2
0.921	1.96	0.964	0.854	2.56	0.914	M_3	3
0.914	2.10	0.955	0.844	2.98	0.909	M_4	4
0.911	2.19	0.949	0.830	3.45	0.900	M_5	5

www.SID.ir

Archive of SID



شکل ۳- نمودارهای مقادیر مشاهداتی و پیش‌بینی شده حاصل از بهترین مدل بیزین در مرحله صحت‌سنجی.

Figure 3. Observed and predicted values of the Bayesian model in test phase.

نتایج مدل‌سازی جریان روزانه رودخانه باراندوزچای با استفاده از ماشین بردار پشتیبان: به‌منظور کاهش دامنه تغییرات داده‌های جریان روزانه رودخانه باراندوزچای و همسان‌سازی اطلاعات ورودی و خروجی، داده‌ها استانداردسازی شد. در مرحله بعد مقادیر بهینه مشخصه‌های مدل SVM شامل ε و C تعیین می‌گردد. 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 شود. بنابراین در حالت کلی برای پیش‌بینی جریان روزانه رودخانه باراندوزچای توسط مدل SVM رگرسیونی، لازم است که مقادیر بهینه سه مشخصه مذکور به‌دست آید که بدین منظور دو مشخصه ε و C توسط الگوریتم بهینه‌سازی جستجوی شبکه[1] و متغیر γ نیز به‌صورت آزمون و خطا محاسبه شد. البته قابل ذکر است که الگوریتم بهینه‌سازی جستجوی شبکه بسیار کند عمل می‌کند و زمان محاسباتی زیادی را به خود اختصاص می‌دهد. برای حل این مشکل در پژوهش مذکور طبق توصیه چن و یو (۲۰۰۷) از برنامه اصلاح‌شده الگوریتم جستجوی شبکه که به نام الگوریتم جستجوی شبکه دو مرحله‌ای[2] معروف است به همراه اعتبارسنجی متقاطع[3] استفاده شد. برای این منظور ابتدا با انتخاب شبکه‌هایی با ابعاد بزرگ محدوده مشخصه‌های ε و C به‌ازای 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 به‌عنوان مقادیر بهینه ε، C و γ انتخاب نمود.

در مطالعه حاضر برای مدل‌سازی جریان روزانه رودخانه باراندوزچای با استفاده از مدل SVM، ۸۰ درصد داده‌ها برای آموزش و ۲۰ درصد برای صحت‌سنجی در نظر گرفته شد. الگوهای ورودی برای مدل SVM نیز به مانند روش شبکه بیزین انتخاب شد تا هر دو روش به‌کار رفته از شرایط یکسانی در اطلاعات دریافتی برخوردار باشند. برای

1- Grid Search

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

www.SID.ir

Archive of SID

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، به‌عنوان 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 مدل‌سازی 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 مدل SVM از دقت بهتری نسبت به روش شبکه بیزین برخوردار است. شکل ۴ نمودارهای پراکندگی و گرافیکی حاصل از مدل SVM برای الگوی M3 نشان می‌دهد.

در شکل ۵ نمودار جعبه‌ای RMSE، برای مدل‌های SVM و BNs در دو حالت آموزش و صحت‌سنجی ارائه شده است. در این شکل خط رسم شده در داخل مستطیل‌های میانه RMSE الگوهای مختلف وردی می‌باشد. عرض متسطیل‌ها در قسمت فوقانی صدک ۷۵ و در قسمت تحتانی صدک ۲۵ را نشان می‌دهد. همچنین قسمت‌های انتهایی خطوط قائم در قسمت فوقانی و تحتانی به‌ترتیب حداکثر و حداقل بیانگر مقدار خطا برای الگوهای مختلف ورودی مدل‌های SVM و BNs می‌باشد. با توجه به شکل ۵ مشاهده می‌شود که عملکرد مدل SVM در دو حال آموزش و تست بسیار بهتر از روش BNs می‌باشد. همچنین در قسمت آموزش مقدار خطای مدل SVM از پراکندگی کم‌تری نسبت به مرحله آموزش BNs برخوردار بوده و این امر در مرحله صحت‌سنجی برعکس شده است. اما با این وجود در مرحله صحت‌سنجی کم‌ترین مقدار خطا به مدل SVM تعلق دارد. نزدیک بودن مقادیر خطا در مرحله صحت‌سنجی مدل BNs بیانگر آن است که از الگوهای M_2 و M_1، زمانی که دسترسی به داده‌های سه روز قبل به هر دلیلی مقدور نباشد یا براساس نظر کارشناسی بایستی پیچیدگی مدل کاهش یابد می‌توان به جای الگوی M_3 استفاده کرد. اما این امکان در روش SVM به‌دلیل پراکندگی بالا در خطاهای الگوهای ورودی مختلف چندان مقدور نمی‌باشد.

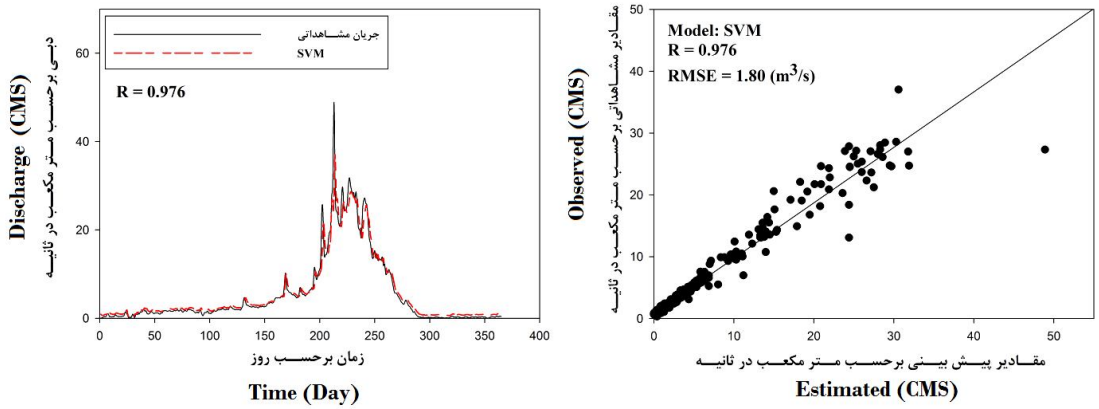
جدول ۵- تحلیل‌های آماری نتایج SVM برای الگوهای مختلف ورودی جریان روزانه.

Table 5. The Statistical analysis of SVM model results for daily river flow inputs.

ردیف (Row)	الگو (Model)	آموزش (Train)			صحت‌سنجی (Train)			متغیرهای مدل SVM (SVM Parameters)		
		R	RMSE (m^3/s)	E	R	RMSE (m^3/s)	E	C	ε	γ
1	M_1	0.935	2.81	0.859	0.975	2.05	0.934	9	0.20	1.500
2	M_2	0.935	2.75	0.864	0.974	1.99	0.938	10	0.30	0.859
3	M_3	0.936	2.77	0.868	0.976	1.80	0.949	10	0.10	0.329
4	M_4	0.934	2.99	0.849	0.975	2.36	0.914	10	0.10	0.270
5	M_5	0.933	2.68	0.850	0.974	2.41	0.931	10	0.40	0.210

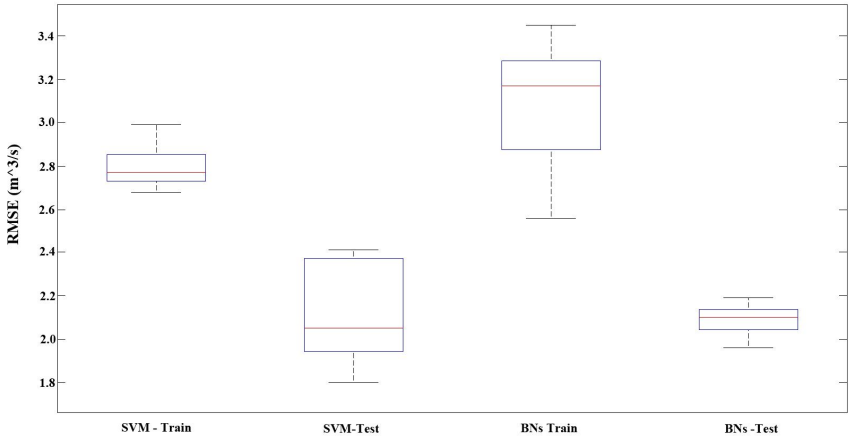
www.SID.ir

Archive of SID



شکل ٤- نمودارهای مقادیر مشاهداتی و مقادیر پیش‌بینی شده حاصل از مدل SVM در مرحله صحت‌سنجی.

Figure 4. Observed and predicted values of the SVM model in test phase.



شکل ٥- نمودار جعبه‌ای RMSE، برای مدل‌های SVM و BNs در دو حالت آموزش و صحت‌سنجی.

Figure 5. The box-plot of the RMSE for SVM and BNs models in train and test phase.

نتیجه‌گیری کلی

تعیین جریان رودخانه‌ها یکی از مؤلفه‌های مهم و تأثیرگذار در مدیریت منابع آب حوضه آبریز می‌باشد. مطالعه این پدیده هیدرولوژیکی کاربردهای فراوانی از جمله در زمینه پیش‌بینی سیلاب دارد. از طرفی به نظر می‌رسد فرآیند جریان رودخانه در مقطع زمانی کوتاه‌مدت به‌ویژه روزانه غیرخطی باشد. منبع اصلی غیرخطی بودن فرآیند جریان‌های کوتاه‌مدت روزانه از غیرخطی بودن فرآیندهای کوتاه‌مدت اقلیمی نظیر بارش و درجه حرارت منشعب شده و غیرخطی بودن واکنش بارش- رواناب نیز باعث پیچیده‌تر شدن فرآیند رودخانه می‌گردد. بنابراین، در این مطالعه جریان روزانه رودخانه باراندوزچای در دوره آماری ۱۳۸۵ تا ۱۳۸۹ با استفاده از روش شبکه بیزین و ماشین بردار پشتیبان مدل‌سازی شد. به‌طورکلی نتایج حاصل از این مطالعه به شرح زیر ارائه می‌شود:

- ارزیابی عملکرد مدل شبکه بیزین بر مبنای دو آماره R و RMSE نشان می‌دهد که دقت مدل‌سازی تا تاخیر سوم رو به بهبود بوده و از آن به بعد کاهش می‌یابد. به‌طوری‌که ضریب همبستگی، میانگین مربعات مجذور خطا و معیار نش- ساتاکلیف در بهترین حالت مدل بیزین به‌ترتیب معادل ۰/۹۶۴، ۱/۹۶ (m^3/s) و ۰/۹۲۱ به‌دست آمد.

www.SID.ir

Archive of SID

- 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 به‌ترتیب معادل ۰/۱۰۰، ۱۰ و ۰/۳۲۹ به‌دست آمد.
- ارزیابی عملکرد SVM بر مبنای دو آماره R و RMSE نشان می‌دهد که دقت مدل‌سازی تا تاخیر سوم رو به بهبود بوده و از آن به بعد کاهش می‌یابد. به‌طوری‌که ضریب همبستگی، میانگین مربعات مجذور خطا و معیار نش- ساتاکلیف در بهترین حالت مدل SVM به‌ترتیب معادل ۰/۹۷۶، ۱/۸۰ (m^3/s) و ۰/۹۴۹ به‌دست آمد.
- در مرحله صحت‌سنجی پراکندگی خطای مدل SVM نسبت به روش شبکه بیزین بیش‌تر بوده اما بهترین دقت متعلق به مدل SVM می‌باشد.
- نتایج مطالعه حاضر در مجموع روش ماشین بردار پشتیبان را به‌عنوان یک روش صریح و دقیق برای پیش‌بینی جریان رودخانه‌ها پیشنهاد می‌کند. این پیشنهاد با نتایج حاصل از مطالعات چوبی (2014)، یو و همکاران (2005) و یو و چن (2006) مطابقت کامل دارد (5, 26, 27).
- با دقت بیش‌تر به شکل‌های ارائه شده برای هر دو روش، می‌توان نتیجه گرفت که عملکرد دو روش SVM و BNs از نظر تحلیل‌های آماری نزدیک به هم بوده، اما از نقطه نظر گرافیکی و تخمین دبی‌های کمینه و بیشینه، مشخص است که هر دو مدل در پیش‌بینی دبی بیشینه خطای بیش‌تری دارند اما در پیش‌بینی دبی متوسط و کمینه عملکرد مدل SVM به مراتب بهتر از مدل BNs می‌باشد.

منابع

1.Adamowski, J., and Prasher, S.O. 2012. Comparison of machine learning methods for runoff forecasting in mountainous watersheds with limited data. J. Water Land Dev. 17: 8. 89-97.

2.Baran, E., and Jantunen, T. 2004. Stakeholder consultation for Bayesian decision support systems in environmental management. J. For. 27: 35. 1-37.

3.Cain, J. 2001. Planning improvement in natural resource management: guideline for using Bayesian networks to support the planning and management of development program in the water sector and beyond. Centre for Ecology and Hydrology (CEH). Wallingford, UK.

4.Chen, S.T., and Yu, P.S. 2007. Real-time probabilistic forecasting of flood stages. J. Hydrol. 340: 63-77.

5.Choubey, V., Mishra, S., and Pandy, S.K. 2014. Time Series Data Mining in Real Time Surface Runoff Forecasting through Support Vector Machine. Inter. J. Com. Appl. 98: 3. 23-30.

6.Dannademehr, A., and Tabatabai, M.R.M. 2010. Prediction of Daily discharge trend of river flow based on genetic programming. J. Water Soil. 24: 2. 325-333. (In Persian)

7.Davies, P. 2007. Bayesian Decision Networks for Management of High Conservation. Report to the Conservation of Freshwater Ecosystem Values Project. Department of Primary Industries and Water, Hobart, Tasmania.

8.Farbodnam, N., Ghorbani, M.A., and Aalami, M.T. 2009. Forecasting river flow using genetic programming (Case study: Lighwan watershed). J. Water Soil Sci. 19: 1. 107-123. (In Persian)

9.Hamel, L.H. 2011. Knowledge discovery with support vector machines (Vol. 3). John Wiley & Sons.

10.Karahan, H., Iplikci, S., Yasar, M., and Gurarslan, G. 2014. River Flow Estimation from Upstream Flow Records Using Support Vector Machines. J. Appl. Math. 1: 1-7.

11.Khalili, K., Fakheri Fard, A., Dinpashoh, Y., and Ghorbani, M.A. 2010. Nonlinearity and testing for stream flow processes by BDS test (Case study: Shaharchai river). J. Water Soil Sci. 22: 2. 25-37. (In Persian)

www.SID.ir

Archive of SID

12.Kuikka, S., and Varis, O. 1997. Uncertainties of climate change impacts in Finnish watersheds: a Bayesian network analysis of expert knowledge. Boreal Environment Research.
13.Madadgar, S., and Moradkhani, H. 2014. Spatio-temporal drought forecasting within Bayesian networks. J. Hydrol. 512: 134-146.
14.McCann, R., Marcot, B., and Ellis, R. 2006. Bayesian Belief Networks: application in ecology and natural resource management. NRC Research Press.
15.Mohajerani, H., Mosaedi, A., Kholghi, M., Meftah Halaghi, M., and Saddodin, A. 2009. Bayesian decision networks introduction and their applications in water resources management. In: Proceedings of 01st national conference on coastal lands, water resources management, 17-18 Nov, Sari Agricultural Sciences and Natural Resources University, Sari, Iran.
16.Mohammadzadeh, J., and Ahmadi, M.R. 2008. Optimization of olive oil extraction to enhance efficiency and quality. J. Agric. Engin. Res. 9: 2. 113-126. (In Persian)
17.Nikbakht Shabazi, A.R. 2008. Application of SVM in predicting the river flow. In: Proceedings of 8th Iranian Hydraulic Conference, 24-26 Nov, Tehran University, Tehran, Iran.
18.Pai, P.F., and Hong, W.C. 2007. A recurrent support vector regression model in rainfall forecasting. Hydrological Process. 21: 819-827.
19.Pollino, C., and Hart, B. 2006. Bayesian network models in natural resource management. Integrated catchment assessment and management (ICAM) Centre of the Australian.
20.Reggiani, P., and Weerts, A. 2008. Bayesian approach to decision-making under uncertainty: An application to real time forecasting in the river Rhine. J. Hydrol. 356: 56-69.
21.Sadoddin, A., Letcher, R.A., Jackeman, A.J., and Newham, L.T.H.A. 2005. Bayesian decision network approach for assessing the ecological impact of salinity management. Mathematics and Computer in Simulation. 69: 162-176.
22.Singh, V.P. 1988. Hydrologic Systems: Rainfall-runoff Modeling. Prentice Hall, Englewood Cliffs, New Jersey.
23.Souza Filho, F.A., and Lall, U. 2003. Seasonal to interannual ensemble streamflow forecasting for Ceara Brazil. J. Water Resour. Res. 39: 11. 1307-1320.
24.Sveinsson Oli, G.B., Salas, J.D., Boes, D.C., and Pielke, R.A. 2003. Modelling the dynamics of long- term variability of hydroclymatic processes. J. Hydrol. 7: 489-501.
25.Vapnik, V.N. 1998. Statistical Learning Theory. Wiley, New York.
26.Yu, P.S., Chen, S.T., and Chang, I.F. 2005. Flood stage forecasting using support vector machines. Geophysical Research Abstracts. 7: 41-76.
27.Yu, P.S., Chen, S.T., and Chang, I.F. 2006. Support vector regression for real-time flood stage forecasting. J. Hydrol. 328: 704-716.

www.SID.ir

Archive of SID

Gorgan University of Agricultural Sciences and Natural Resources

J. of Water and Soil Conservation, Vol. 22(6), 2016
http://jwsc.gau.ac.ir

Comparing the performance of Support Vector Machines and Bayesian networks in predicting daily river flow (Case study: Baranduz Chai River)

***F. Ahmadi[1], F. Radmanesh[2] and R. Mirabbasi Najafabadi[3]**

[1]Ph.D. Student, Dept. of Water Resources Engineering, Shahid Chamran University of Ahvaz,
[2]Associate Prof., Dept. of Water Resources Engineering, Shahid Chamran University of Ahvaz,
[3]Assistant Prof., Dept. of Water Engineering, Shahrekord University
Received: 10/21/2014; Accepted: 06/01/2015

Abstract

Background and Objectives: Since the river flow forecasting is an important issue in water resources engineering, the researchers have considered and applied different methods in this field. In recent decades the artificial intelligence (AI) methods emerge as a powerful tool for detecting the relationship among different variables of hydrological systems. Support Vector Machine and Bayesian networks are two types of artificial intelligence methods which are widely used in recent years for river flow forecasting.

Materials and Methods: In this study, Support Vector Machines (SVM) and Bayesian networks (BNs) methods were used to forecast daily discharge of the Baranduz Chai River. The support vector machine was a novel neural network algorithm and originally applied to pattern recognition problems and regression. In recent years, decision support systems (DSS) have been developed to integrate the best available knowledge for making appropriate decision. Among the various kinds of DSS, Bayesian networks have proven quite useful since this modeling approach can integrate quantitative information and data as well as qualitative expert knowledge.

Results: The daily discharge data of the Baranduz Chai River measured at the Dizaj hydrometric station during 2006 to 2010 were used for modeling, which 80% of the data were used for training and remaining 20% used for testing of models. In order to model river flow by SVM and BNs methods, the river flow discharge of over 5 days with daily time steps were used and the results evaluated based on root mean square error (RMSE), correlation coefficient (R) and Nash-Sutcllife (NS) criteria. The results showed that at first, the accuracy of models increased with increasing of memory, as the most accuracy obtained for combination #3 of input variables in both of the methods. Comparing the performance of SVM and BNs models indicated that the accuracy of the SVM method with the R=0.976, RMSE=1.80 (m^3/s) and NS=0.949 was slightly more than BNs method with R=0.964, RMSE=1.96 (m^3/s) and NS=0.921.

Conclusion: According to obtained results, it can be concluded that the performance of SVM and BNs models are close together, but both of considered models had a remarkable error in estimating high discharge values. However, the SVM model had better performance than BNs in forecasting average and the minimum discharge values.

Keywords: Bayesian networks (BNs), Support Vector Machines (SVM), Modeling

* Corresponding Author; Email: farsahad.paper@yahoo.com

www.SID.ir