

ارزیابی توانایی یک روش آموزش ماشین در تخمین حداکثر دبی سیلاب ناشی از شکست سد

حامد فرهادی^۱، کاظم اسماعیلی^۲

چکیده

در این مقاله توانایی روش آموزش ماشین Support Vector Machine (SVM) در پیش بینی حداکثر دبی خروجی سیلاب ناشی از شکست سدهای خاکی بررسی شده است. پارامترهای ورودی مدل مورد نظر، دو پارامتر مخزن در زمان شکست یعنی ارتفاع آب و حجم آب پشت سد انتخاب شد که برای آموزش این مدل‌ها از داده‌های جمع آوری شده در منابع مختلف استفاده شده است. از مجموع ۱۱۲ داده، ۷۰ درصد آن جهت آموزش مدل‌ها و ۳۰ درصد آن جهت صحت سنجی، به نحوی که این دو زیرمجموعه از لحاظ آماری اختلاف معنی داری نداشته باشند، انتخاب شد. بعد از بررسی چهار مدل SVM، مشخص شد که استفاده از کرنل تابع پایه شعاعی بهترین نتیجه را در تخمین این پدیده حاصل می‌دهد (نتایج آماری این روش در تخمین پدیده مورد نظر با $R^2=0/96$ ، $RMSE=0/03$ و $R^2=0/94$ و $RMSE=0/05$ به ترتیب در فاز آموزش و آزمون). مقایسه‌ای نیز بین ۶ رابطه تجربی کلاسیک و مدل توصیه شده صورت گرفت که نتایج نشانگر عملکرد ضعیف روابط تجربی در مقایسه با مدل پیشنهادی است. در نهایت با توجه به اهمیت مدیریت هنگام شکست سد استفاده از روش آموزش ماشین پیشنهادی برای تخمین مقدار ماکسیمم دبی خروجی از سد پیشنهاد می‌شود.

واژه‌های کلیدی: آموزش ماشین، دبی سیلاب، سد خاکی، شکست سد، ماشین بردار

پشتیبان، هوش مصنوعی

مقدمه

شکست، تحقیقات زیادی در زمینه آنالیز شکست و تعیین هیدروگراف خروجی آن صورت گرفته است. پارامترهای شکست سد را می‌توان به دو گروه هندسی و هیدرولیکی تقسیم کرد. مقطع شکست معمولاً دوزنقه‌ای شکل است که با ارتباط پارامترهای هندسی می‌توان اندازه و شکل مقطع شکست را بدست آورد. پارامترهای هیدرولیکی شامل نرخ ماکسیمم سیلاب خروجی و زمان شکست می‌باشد. بعد از آغاز شکست نرخ خروج آب با مرور زمان افزایش پیدا می‌کند و به بیشترین مقدار خود می‌رسد و در ادامه با کاهش حجم آب سد نرخ سیلاب خروجی نیز کاهش می‌یابد و تا زمانی که هیچ آبی در مخزن سد باقی نماند ادامه پیدا

سدها برای بهبود شرایط کیفی زندگی بشر، با هدف تأمین آب شرب و کشاورزی، تبدیل انرژی و کنترل سیلاب، ساخته می‌شوند اما به دلیل بوجود آمدن مخزنی با حجم عظیم آب در بالادست سد و ریسک زیاد مالی و جانی رها شدن ناگهانی آن به پایین دست، آنالیز شکست سد را در طراحی سدهای بزرگ الزام آور می‌کند. شکست سدهای خاکی از جمله وقایعی است که علاوه بر خسارات شدید مالی، موجب از دست رفتن جان انسان‌ها و همچنین پیامدهای زیست محیطی نظیر آب‌گرفتگی در سطح وسیع و انتقال رسوب می‌شود. از این‌رو در جهت کنترل حوادث ناشی از

^۱ دانشجوی دکتری سازه‌های آبی، گروه علوم و مهندسی آب دانشگاه فردوسی مشهد، ایران، ۰۹۱۳۳۹۸۷۹۳۹. farhadi.edu@gmail.com

^۲ عضو هیئت علمی گروه علوم و مهندسی آب دانشگاه فردوسی مشهد، ایران، ۰۵۱۳۸۸۰۵۸۱۳. esmaili@um.ac.ir (نویسنده مسئول)

می‌کند. مطالعاتی که پیرامون شکست سد انجام شده است را می‌توان به ۳ گروه تقسیم کرد؛ مدل‌های فیزیکی که در مقیاس مشخص در قالب مدل‌های آزمایشگاهی، شکست سد را مورد مطالعه قرار می‌دهند که در این زمینه می‌توان از مطالعات Coleman et al. (2002) و Gaucher et al. (2010) نام برد. این مدل‌ها بدلیل محدودیت اجرایی و مشکلات ناشی از اثرات مقیاس بیشتر برای اعتبار سنجی مدل‌های عددی و تجربی قابل استفاده هستند. مدل‌های عددی از تکنیک‌های حل عددی، که در اغلب موارد حل معادلات سنت و نانت است، استفاده می‌کند در این رابطه می‌توان مطالعات Tsai, Ponce et al. (2003) و Xia (2010) نام برد. علیرغم مطالعات و تحقیقات زیاد در زمینه کاربرد مدل‌های عددی، آن‌ها به طور کامل قابل اتکا نیستند و تنها می‌تواند به عنوان یک راهنما برای آشنایی بیشتر با پدیده مورد استفاده قرار گیرند (Carling et al., 2009). در مدل‌ها و روابط تجربی فیزیک مسئله چندان نقشی ندارد و بر اساس مشاهدات یا رابطه بین داده‌ها می‌باشند. از مدل‌های تجربی می‌توان به روابط رگرسیونی اشاره کرد که محققین زیادی برای پیش‌بینی حداکثر دبی سیلاب از این روابط استفاده کردند (SCS, 1981; USBR, 1982; Macdonald and Langridge-Monopolis, 1984; Evans, 1986; Froehlich, 1995; Pierce et al., 2009; Xia et al., 2010). در جدول ۱ برخی روابط که کاربرد بیشتری دارند درج شده است. روابط مذکور نشان می‌دهد محققین ارتباط حداکثر دبی سیلاب خروجی را با پارامترهای مخزن یعنی H_w (رابطه USBR, 1982)، V_w (رابطه Evans, 1986)، $(V_w.H_w)$ (رابطه Macdonald and Langridge-Monopolis, 1984)، که به آن فاکتور سد می‌گویند (Pierce et al., 2009) و ترکیب دو پارامتر H_w و V_w

(روابط Froehlich, 1995 و Pierce et al., 2009) را مورد بررسی قرار دادند. این روش‌ها به دلیل استفاده از داده‌های محدود جهت مطالعه و همچنین تنها با استفاده از روابط کلاسیک رگرسیونی قابل تعمیم در شرایط متفاوت نمی‌باشند. یک راهکار دیگر بهره گرفتن از روش‌های آموزش ماشین است که یکی از کاربردهای آن توانایی زیاد در یافتن روابط غیرخطی بین داده‌هاست. به همین دلیل، این روش‌ها توجه بسیاری از محققین در زمینه‌های گوناگون را به خود جلب کرده است. اخیراً این روش‌ها نتایج بسیار قابل قبولی را در زمینه موضوعات مهندسی منابع آب از خود نشان داده‌اند، چنانکه، روش شبکه عصبی مصنوعی (ANN)^۱ در تخمین بارندگی (Smith and Eli, 1995; Tokar and Johnson, 1999; Dawson and Wilby, 2001) پیش‌بینی جریان (Chane et al., 2002; Chiang et al., 2004; Johnson et al., 1994) و مدلسازی آب‌های زیرزمینی (Ranjithan et al., 1993; al., 1995)، روش ماشین بردار پشتیبان (SVM)^۲ در مدلسازی بارندگی-رواناب (Dibike et al., 2001; Bray and Han, 2004) پیش‌بینی سیلاب (Chen and Yu, Han et al., 2007) روش برنامه نویسی ژنتیک (GP)^۳ در مدلسازی بارندگی و رواناب (Liong et al., 2002; Whigham and Crapper, 2001)، پیش‌بینی جریان (Sivapragasam et al., 2008) و مطالعات هیدرولوژی (Makkeasorn et al., 2008; Wang et al., 2009)، روش ANFIS^۴ در پیش‌بینی جریان (Chang and Chang, 2001)، مدلسازی آب‌های زیرزمینی (Dixon, 2005)، مطالعات هیدرولوژی (Chang and Chang, 2001; Nayak et al., 2004) مورد استفاده قرار گرفته‌اند.

1- Artificial Neural Network
2- Support Vector Machine
3- Genetic Programming
4- Adaptive Neuro-Fuzzy Inference System

جدول (۱): نمونه روابط تجربی ارائه شده برای تخمین دبی سیلاب خروجی حداکثر

رابطه	محقق
$Q_p = 19.1H_w^{1.85}$	USBR (1982)
$Q_p = 0.72V_w^{0.53}$	Evans (1986)
$Q_p = 1.154(V_w \cdot H_w)^{0.412}$	MacDonald and Langridge-Monopolis (1984)
$Q_p = 0.607(H_w^{1.24} V_w^{0.295})$	Froehlich (1995)
$Q_p = 0.038(H_w^{1.09} V_w^{0.475})$	Pierce et al (2009)

مواد و روش‌ها

روش ماشین بردار پشتیبان (SVM)

روش ماشین بردار پشتیبان (SVM) یک روش آموزش ماشین است که توسط Vapnik و همکارانش ارائه شد (Cortes and Vapnik, 1995; Vapnik, 1995). کاربرد اصلی روش SVM ابتدا در طبقه‌بندی و تشخیص الگو معرفی شد که با روش‌های دیگر رقابت می‌کرد (Schölkopf et al., 1995). رگرسیون، کاربردی از این روش است که در مطالعات (Smola (1996) ارائه شد و نتایج قابل توجه آن در رگرسیون و پیش‌بینی سری‌های زمانی توسط Mattera and Hayek (1999) گزارش شد. در این روش، رگرسیون به دو صورت انجام می‌شود که تحت نام‌های v -SVM و ϵ -SVM شناخته می‌شوند (Smola, 1996). که در تحقیق حاضر از ϵ -SVM برای رگرسیون استفاده شده است. برای توضیح روش، یک سری داده‌ها برای آموزش مدل به صورت $\{(x_1, y_1), (x_2, y_2), \dots, (x_l, y_l)\} \subset \chi \times R$ در نظر گرفته شود که در آن l و χ تعداد و بُعد داده‌های ورودی هستند. تابع خطی در فضای یک بُعدی می‌تواند به صورت رابطه ۱ تعریف شود،

$$f(x) = w \cdot x + b \quad (1)$$

w بردار وزنی و b اریب یا مقدار فاصله از مبدأ می‌باشد که از طریق مینیمم کردن تابع ریسک که در رابطه ۲ نشان داده شده است به دست می‌آیند،

از آنجایی که فیزیک مسئله شکست سد و مدل‌سازی سیلاب حاصل از آن در پایین دست بسیار پیچیده است، روش‌های آموزش ماشین می‌توانند به عنوان روش جایگزین در حل مسئله مؤثر باشند. Amini et al. (2011) و Hooshyaripor and Babaeyan al. (2012) و Tahershamsi et al. (2014) و Hooshyaripor et al. (2014) با آموزش شبکه عصبی مصنوعی (ANN) و Sattar (2015)، با استفاده از روش برنامه نویسی بیان ژن (GEP)^۱ با در نظر گرفتن داده‌های تاریخی به وقوع پیوسته و داده‌های آزمایشگاهی، حداکثر دبی سیلاب حاصل از شکست سد را مدل‌سازی کردند. همچنین فرهادی و زمردیان (۱۳۹۱) با خروجی حاصله از مدل مبتنی بر فیزیک BREACH و با استفاده از روش SVM پارامترهای شکست سد از جمله حداکثر دبی سیلاب حاصله را تخمین زدند.

هدف از این مقاله بررسی توانمندی یک روش آموزش ماشین (SVM) در پیش‌بینی و تخمین حداکثر دبی سیلاب ناشی از شکست سد می‌باشد. تعداد ۱۱۲ داده موجود در مقالات و مستندات علمی جهت آموزش و آزمون این روش مورد استفاده قرار گرفت. سپس این مدل با روابط تجربی کلاسیک در تخمین حداکثر دبی خروجی از سد هنگام شکست مقایسه شد. در ادامه ضمن توضیح مختصر در مورد تئوری مدل، نتایج حاصله از روش‌ها مورد بررسی قرار گرفت و در انتها نتیجه‌گیری کلی ذکر شده است.

$$R = C \frac{1}{l} \sum_{i=1}^N L_{\varepsilon}(y_i, f_i) + \frac{1}{2} \|w\|^2 \quad (2)$$

که C یک پارامتر کنترل است که از سعی و خطا بین پیچیدگی مدل و خطای آموزش تعیین می شود. L_{ε} تابع عدم حساسیت که در رابطه ۳ تعریف شده است و ε ، انحراف مقادیر واقعی y_i از تخمین زده شده f_i می باشد.

$$L_{\varepsilon}(x_i, y_i, f(x_i)) = \begin{cases} 0 & \text{if } |y_i - f(x_i)| \leq \varepsilon \\ |y_i - f(x_i)| - \varepsilon & \text{otherwise} \end{cases} \quad (3)$$

داده های در بازه $|y_i - f(x_i)| \leq \varepsilon$ قابل صرف نظر است و داده های خارج از محدوده ε و روی خط جداکننده، بردارهای پشتیبان نامیده می شوند، و تنها بردارهای پشتیبان در حل مسئله بهینه یابی استفاده می شوند. از نگاه کاربرد رگرسیون مدل SVM، هدف پیدا کردن $f(x)$ است که بیشترین اختلاف ε با مقادیر واقعی هدف y_i را داشته باشد و در عین حال در حد امکان مدل پیچیده نباشد (Smola, 1996). بیشترین اختلاف در محدوده ε ، برای جلوگیری از مسئله overfitting در پیش بینی داده های دیده نشده توسط مدل است که با این رویکرد توانایی مدل در

پیش بینی بیشتر خواهد بود. می توان مانند یک تیوب یا یک محدوده در اطراف خط (در فضای یک بعدی) یا ابر صفحه (در فضای با بعد بالاتر) جداکننده داده ها در نظر گرفت که نقاط خارج از این محدوده خطای آموزش (ξ_i, ξ_i^*) قلمداد می شود (شکل ۱). بنابراین با لحاظ کردن به حداقل رساندن پیچیدگی مدل نیاز است این خطاهای آموزش نیز مینیمم شوند. در نهایت

مسئله بهینه یابی همانند رابطه ۴ می شود.

$$\text{minimize} \left(\frac{1}{2} \|w\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*) \right) \quad (4)$$

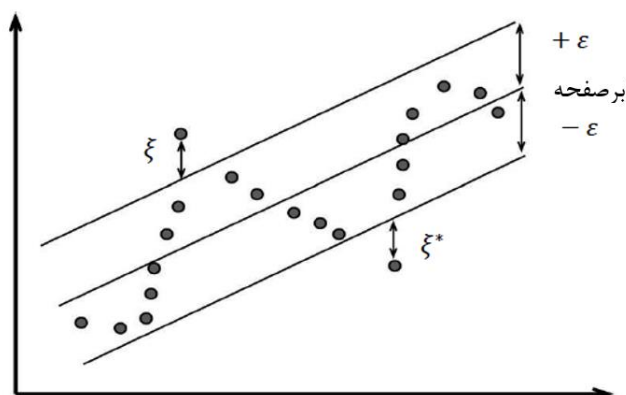
تحت قیود،

$$y_i - (w \cdot x) - b \leq \varepsilon + \xi_i$$

$$(w \cdot x) + b - y_i \leq \varepsilon + \xi_i^*$$

$$\xi_i, \xi_i^* \geq 0$$

برای حل این مسئله بهینه یابی، از روش برنامه ریزی درجه دوم^۱ استفاده می شود. که در نهایت بعد از تشکیل تابع لاگرانژ، که شامل تابع هدف و قیود می باشد (رابطه ۵)، و مشتق گیری از تابع لاگرانژ نسبت به پارامترهای w ، b ، ξ_i و ξ_i^* ، به رابطه نهایی دست خواهیم یافت (رابطه ۶).



شکل (۱): شکل شماتیک نشان دهنده خطای آموزش مثبت و منفی

پیش فرآیند داده‌ها برای ساختن مدل

از مجموع ۱۱۲ داده جمع آوری شده از مقالات و گزارش‌ها (SCS, 1981; Singh and Scarlatos, 1988; Wahl, 1988; Yumei, 1994; Fujia and Burge, 2004; Vaskin et al., 2004; Hanson et al., 2005; Xu and Zhang, 2009; FERC, 2006)، ۷۰ درصد آن به آموزش مدل‌ها و ۳۰ درصد به صحت سنجی اختصاص می‌یابد. محدوده داده‌های استفاده شده در این مطالعه در جدول ۲ قابل مشاهده است. انتخاب داده‌های مربوط به آموزش و آزمون مدل به نحوی صورت گرفت که دو زیرمجموعه آموزش و آزمون از لحاظ آماری شبیه یکدیگر باشند. بدین صورت که با آزمون‌های t و F ، بررسی شد که به ترتیب میانگین و انحراف معیار این دو دسته تفاوت آماری معنی داری در سطح اطمینان ۹۵ درصد نداشته باشند. همچنین ماکسیمم و مینیمم پارامترها در محدوده آموزش قرار گرفت که مشکل برون‌یابی و تعمیم این مدل برای داده‌های با دامنه بیشتر پیش نیاید. قبل از اعمال روش‌ها، داده‌ها به دلیل متفاوت بودن جنس و محدوده متغیرها توسط رابطه ۱۱ بین (۰ و ۱) نرمال شدند. در این تحقیق، از پارامترهای H_w و V_w به عنوان ورودی به مدل‌ها استفاده شد (رابطه ۱۲).

$$x_n = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (11)$$

$$Q_p = f(H_w, V_w) \quad (12)$$

$$L = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*) - \sum_{i=1}^n \alpha_i [(w_i \cdot x + b) - y_i + (\varepsilon + \xi_i)] - \sum_{i=1}^n \alpha_i^* [y_i - (w_i \cdot x + b) + (\varepsilon + \xi_i^*)] - \sum_{i=1}^n [r_i \xi_i + r_i^* \xi_i^*] \quad (5)$$

$f(x_i) = \sum (\alpha_i^* - \alpha_i) K(x_i, x_j) + b \quad (6)$
که $\alpha_i^*, \alpha_i \geq 0$ ضرایب لاگرانژی، b ترم اریب یا عرض از مبدأ و $K(x_i, x)$ تابع کرنل^۱ می‌باشد. کرنل‌ها توابعی هستند که داده‌های با قابلیت جداپذیری غیرخطی را به بُعدی بالاتر نگاشت می‌کند تا در آن بُعد، داده‌ها به طور خطی طبقه بندی شوند (تا عمل رگرسیون به صورت خطی میسر شود). در این تحقیق چهار تابع کرنل خطی^۲، چندجمله‌ای^۳، تابع پایه شعاعی (RBF)^۴، سیگموئید^۵، به ترتیب نشان داده شده در روابط ۷ تا ۱۰، برای مدل SVM در نظر گرفته شد و در بهترین نتایج با یکدیگر مقایسه شدند.

$$K(X_i, X_j) = X_i \cdot X_j \quad (7)$$

$$K(X_i, X_j) = \gamma (X_i \cdot X_j + Coef)^d \quad (8)$$

$$K(X_i, X_j) = -\gamma \|X_i - X_j\|^2 \quad (9)$$

$$K(X_i, X_j) = \tanh(\gamma X_i \cdot X_j + Coef) \quad (10)$$

که در آن‌ها d و $Coef$ عدد ثابت هستند که همانند پارامترهای γ ، C و ε مجهول می‌باشند.

جدول (۲): محدوده داده‌های جمع آوری شده از مستندات علمی

پارامتر	محدوده
$H_w (m)$	۱/۷۷-۳۷/۴
$V_w (m^3)$	۰/۹۴-۰۰۳۷
$Q_p (m^3/s)$	۰/۷۸۱۰۰-۲۸

2- Kernel
3- Linear
1- Polynomial
2- Radial Basis Function
3- Sigmoid

ابزار تحلیل آماری نتایج

تحلیل نتایج تنها با یک پارامتر آماری، به دلیل محدودیت‌های خاص هر پارامتر آماری، نمی‌تواند با قطعیت مناسب همراه باشد. بنابراین در این مطالعه از پارامترهای آماری ضریب تبیین (R^2)، جذر میانگین مربعات خطا (RMSE) و خطای باقیمانده (Residual) برای ارزیابی نتایج مدل SVM بهره گرفته شد. روابط مربوط به پارامترهای آمای ذکر شده در ذیل آمده است.

$$R^2 = 1 - \frac{SS_E}{SS_T} = 1 - \frac{\sum_1^n (y_i - f_i)^2}{\sum_1^n (y_i - \bar{y})^2} \quad (13)$$

$$RMSE = \sqrt{\frac{\sum (y_i - f_i)^2}{n}} \quad (14)$$

$$Residual = y_i - f_i \quad (15)$$

که در آن SS_T و SS_E مجموع مربعات خطا و کل مربعات خطای داده‌ها هستند. y_i ، f_i ، $\bar{y}$ و n به ترتیب، داده‌های مشاهده‌ای، داده‌های پیش‌بینی شده توسط مدل‌ها، میانگین داده‌های مشاهده‌ای و تعداد داده‌ها می‌باشند.

تحلیل نتایج

برای دستیابی به نتیجه مناسب چهار مدل SVM با کرنل‌های خطی، چندجمله‌ای، پایه شعاعی و

سیگموئید مورد مطالعه قرار گرفت. برای هر چهارمدل SVM پارامترهای بهینه که بهترین نتیجه را براساس پارامتر آماری R^2 نتیجه می‌داد با سعی و خطا بدست آمد. پارامترهای مناسب استخراج شده برای پیش‌بینی در جدول ۳ قابل مشاهده است، همانطور که از نتایج قابل استنباط است، SVM با کرنل پایه شعاعی با سه پارامتر $C=10$ ، $\gamma = 10$ ، $\varepsilon = 0.0096$ بهترین نتیجه را حاصل داده است.

با در نظر گرفتن مشابهت آماری بین دو زیرمجموعه آموزش و آزمون از خطر overfitting (که به معنی نتیجه خوب برای آموزش و عدم توانایی در ارائه نتیجه مناسب در دامنه داده خارج از داده‌های آموزش است) اجتناب شد. که در نتایج بدست آمده در جدول ۳ قابل استنباط است که نشان داده شد با این روش می‌توان به هدف مورد نظر (اجتناب از overfitting) دست یافت. این اشکال که با در نظر گرفتن دو زیر مجموعه با تشابه آماری معنی دار محیط بسته‌ای برای ارائه نتایج فراهم شده است و در صورت آموزش درست مدل بدون شک به نتیجه مناسبی برای داده‌های آزمون خواهیم رسید، با در نظر گرفتن ماکسیمم و مینیمم داده‌ها در دامنه آموزش، توانایی مدل در ارائه نتیجه مناسب در دامنه وسیع‌تری تضمین داده شده است.

جدول (۳): نتایج حاصل از استفاده از مدل SVM با توابع کرنل گوناگون

نوع تابع کرنل	فاز آموزش R^2	فاز آزمون R^2	کل داده‌ها R^2	پارامترهای بهینه
خطی	۰/۷۵	۰/۹۵	۰/۸۲	$C=10$, $\varepsilon=0$
چند جمله‌ای	۰/۵۱	۰/۹۴	۰/۷۰	$=3$, $\gamma=10$, d Coef=0
تابع پایه شعاعی	۰/۹۶	۰/۹۴	۰/۹۴	$\gamma = 10$, $C=10$ $\varepsilon = 0.0096$
سیگموئید	۰/۵۸	۰/۴۴	۰/۵۳	$=10\gamma$ Coef=0

در جدول ۴ و ۵ نتایج حاصل از مدل‌های پیشنهاد شده با روابط تجربی کلاسیک مورد مقایسه قرار گرفت.

نتایج نشان می‌دهد که این روابط در تخمین حداکثر دبی خروجی سیلاب در داده‌های جمع آوری شده

ضعیف عمل می‌کنند. اما در مقابل، مدل پیشنهاد شده SVM با نتایج مناسب توانایی خود را در تخمین این پارامتر نشان داد. پیشنهاد میشود که مدل SVM با تابع پایه شعاعی با پارامترهای معرفی شده در پیش‌بینی دبی ماکسیمم سیلاب حاصله از شکست سدهای خاکی مورد استفاده قرار گیرد.

جدول (۴): نتایج کلی مدل‌های استفاده شده و مقایسه آن با روابط تجربی

روش	مرحله آموزش			مرحله صحت سنجی			تمام داده‌ها		
	متوسط	خطای	R ²	متوسط	خطای	R ²	متوسط	خطای	R ²
	باقیمانده			باقیمانده			باقیمانده		
SVM	-۰/۰۰۱	۰/۰۳۳	۰/۹۶	-۰/۰۰۰۷	۰/۰۵۴	۰/۹۴	۰/۰۰۱	۰/۰۳۳	۰/۹۴

جدول (۵): مقایسه نتایج روش SVM با روش‌های کلاسیک

روش	متوسط خطای باقیمانده	RMSE	R ²
SVM	۰/۰۰۱	۰/۰۳۳	۰/۹۴
USBR (1982)	-۱/۰۵	۲/۶۲	۰/۳۱
Evans (1986)	۰/۰۸۸	۰/۱۰	۰/۶۲
Froehlich (1995)	۰/۱۴	۰/۱۸۰	۰/۵۸
MacDonald & Langridge-Monopolis (1984)	۰/۰۶۸	۰/۱۰۴	۰/۶۸
Pierce (2009)	۰/۱۶۵	۰/۱۷	۰/۶۵

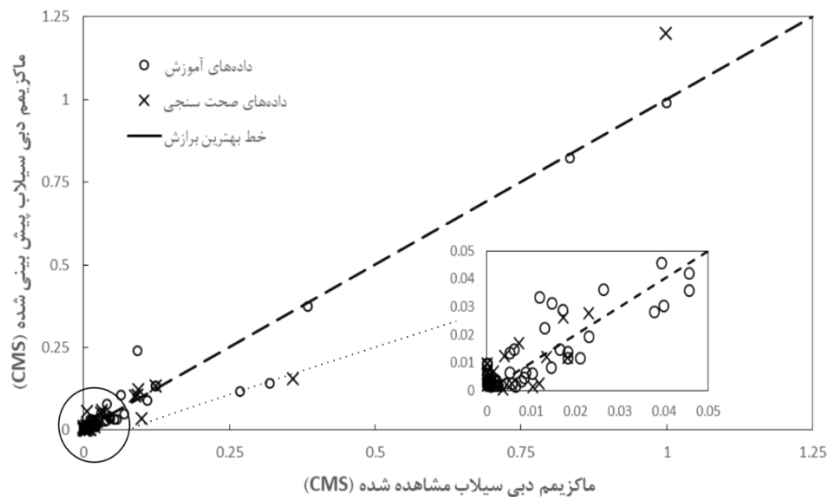
نتیجه گیری

در این مطالعه توانایی مدل SVM در تخمین حداکثر دبی سیلاب حاصل شکست ۱۱۲ مورد سد مورد ارزیابی قرار گرفت. مدل‌های تجربی کلاسیک از روابط رگرسیونی ساده برای تخمین این پارامتر استفاده می‌کنند، مدل‌های آزمایشگاهی نیز مشکل پیاده کردن شرایط طبیعی در مقیاس کوچک (اثر مقیاس) دارد و همچنین عدم قطعیت در برداشت داده‌ها از مشکلات این نوع مدل‌سازی است. روش‌های عددی نیز به داده‌های فراوان جهت شبیه‌سازی این پدیده نیازمند است. بنابراین استفاده از روش‌هایی که فارغ از فیزیک پدیده، با توانایی بالا، قابلیت کشف رابطه غیرخطی بین داده‌ها را دارند (روش‌های آموزش ماشین)، می‌تواند

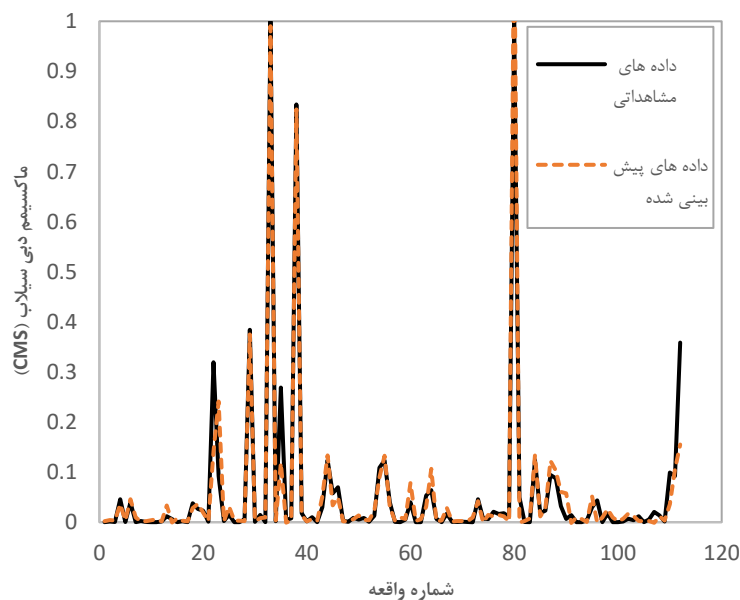
امتیاز ویژه‌ای باشد. شرایط محدود کننده روش‌های مذکور در مورد این مدل‌ها صدق نمی‌کند. اما عیب این روش‌ها در توانایی تخمین پارامتر در محدوده داده‌هاییست که مدل با آن آموزش دیده است. بنابراین در این مطالعه سعی شد که از داده‌های متفاوت چاپ شده در اوراق علمی برای آموزش مدل SVM استفاده شود. در انتخاب داده‌ها به نحوی عمل شد که مجموعه داده‌های آموزش و آزمون از لحاظ آماری تفاوت معنی داری نداشته باشند که این کار با آزمون‌های t و F انجام شد. چهار تابع کرنل خطی، چندجمله‌ای، تابع پایه شعاعی، سیگموئید مورد بررسی قرار گرفتند که در نهایت مدل SVM با پارامترهای $C=10$ ، $\gamma=10$ و $\epsilon=0.0096$ مربوط به تابع کرنل پایه شعاعی به

عنوان راهکار پیشنهادی به عنوان یک روش آموزش ماشین پیشنهاد شد. در ادامه نتایج این مدل‌ها با نتایج عملکرد روابط تجربی کلاسیک، که بیشتر مورد استفاده محققین قرار می‌گیرد، مقایسه شد. در این مقایسه توانایی بالای مدل پیشنهادی نسبت به روابط

تجربی کلاسیک مشهود بود. پیشنهاد می‌شود که مدل SVM با ساختار معرفی شده برای تخمین حداکثر دبی خروجی حاصل از شکست سد مورد استفاده قرار گیرد.



شکل (۲): نمودار ماکسیمم دبی سیلاب مشاهده شده در مقابل پیش بینی شده توسط مدل SVM



شکل (۳): مقایسه ماکسیمم دبی سیلاب مشاهده شده و پیش بینی شده توسط مدل SVM در شکست سدهای مورد مطالعه

منابع

- فرهادی، ح.، م. زمردیان. ۱۳۹۱. برآورد ماکسیمم دبی شکست سدهای خاکی و زمان شکست با استفاده از روش SVM، یازدهمین کنفرانس هیدرولیک ایران، ارومیه.
- Babaeyan-Amini, A., V. Nourani and H. Hakimzadeh. 2011. Application of artificial intelligence tools to estimate peak outflow from earth dam breach. *International Journal of Earth Sciences and Engineering*, 4: 243-246.
- Bray, M. and D. Han. 2004. Identification of support vector machines for runoff modelling. *Journal of Hydroinformatics*, 6: 265-280.
- Burge, T. 2004. Big Bay Dam: Evaluation of failure. Land Partners Limited Partnership, Hattiesburg, Miss.
- Carling, P., I. Villanueva, J. Herget, N. Wright, P. Borodavko and H. Morvan. 2010. Unsteady 1D and 2D hydraulic models with ice dam break for Quaternary megaflood, Altai Mountains, southern Siberia. *Global and Planetary Change*, 70: 24-34.
- Chang, F., L. C. Chang and H. L. Huang. 2002. Real-time recurrent learning neural network for stream-flow forecasting. *Hydrological Processes*, 16: 2577-2588.
- Chang, L. C. and F. J. Chang. 2001. Intelligent control for modelling of real-time reservoir operation. *Hydrological processes*, 15: 1621-1634.
- Chen, S. T. and P. S. Yu. 2007. Pruning of support vector networks on flood forecasting. *Journal of Hydrology*, 347: 67-78.
- Chiang, Y. M., L. C. Chang, and F. J. Chang. 2004. Comparison of static-feedforward and dynamic-feedback neural networks for rainfall-runoff modeling. *Journal of hydrology*, 290: 297-311.
- Coleman, S. E., D. P. Andrews and M. G. Webby. 2002. Overtopping breaching of noncohesive homogeneous embankments. *Journal of Hydraulic Engineering*, 128: 829-838.
- Cortes, C. and V. Vapnik. 1995. Support-vector networks. *Machine learning*, 20: 273-297.
- Dawson, C. and R. Wilby. 2001. Hydrological modelling using artificial neural networks. *Progress in physical Geography*, 25: 80-108.
- Dibike, Y. B., S. Velickov, D. Solomatine, and M. B. Abbott. 2001. Model induction with support vector machines: introduction and applications. *Journal of Computing in Civil Engineering*, 15: 208-216.
- Dixon, B. 2005. Groundwater vulnerability mapping: a GIS and fuzzy rule based integrated tool. *Applied Geography*, 25: 327-347.
- Evans, S. G. 1986. The maximum discharge of outburst floods caused by the breaching of man-made and natural dams. *Canadian Geotechnical Journal*, 23: 385-387.
- FERC. 2006. Report of findings on the overtopping and embankment breach of the upper dam – Taum Sauk Pumped Storage Project. FERC No. 2277, April, 239 p.
- Froehlich, D. C. 1995. Peak outflow from breached embankment dam. *Journal of Water Resources Planning and Management*, 121: 90-97.

Gaucher, J., C. Marche and T. F. Mahdi. 2010. Experimental investigation of the hydraulic erosion of noncohesive compacted soils. *Journal of Hydraulic Engineering*, 136: 901-913.

Han, D., L. Chan and N. Zhu. 2007. Flood forecasting using support vector machines. *Journal of hydroinformatics*, 9: 267-276.

Hanson, G., K. Cook and S. Hunt. 2005. Physical modeling of overtopping erosion and breach formation of cohesive embankments. *Transactions of the ASAE*, 48: 1783-1794.

Hassan, M., M. Morris, G. Hanson and K. Lakhal. 2004. Breach formation: Laboratory and numerical modeling of breach formation. Association of State Dam Safety Officials: Dam Safety.

Hooshyaripor, F. and A. Tahershamsi. 2012. Comparing the performance of neural networks for predicting peak outflow from breached embankments when back propagation algorithms meet evolutionary algorithms. *International Journal of Hydraulic Engineering*, 1: 55-67.

Hooshyaripor, F., A. Tahershamsi, and S. Golian. 2014. Application of copula method and neural networks for predicting peak outflow from breached embankments. *Journal of Hydro-environment Research*, 8: 292-303.

Johnson, V. M. and L. L. Rogers. 1995. Location Analysis in Ground-Water Remediation Using Neural Networks. *Ground Water*, 33: 749-758.

Karunanithi, N., W. J. Grenney, D. Whitley and K. Bovee. 1994. Neural networks for river flow prediction. *Journal of Computing in Civil Engineering*, 8: 201-220.

Liong, S. Y., T. R. Gautam, S. T. Khu, V. Babovic, M. Keijzer and N. Muttill. 2002. Genetic programming: A new paradigm in rainfall runoff modeling¹. *Journal of the American Water Resources Association*, 38: 705-718.

Macdonald, T. C. and J. Langridge-Monopolis. 1984. Breaching characteristics of dam failures. *Journal of Hydraulic Engineering*, 110: 567-586.

Makkeasorn, A., N. B. Chang and X. Zhou. 2008. Short-term streamflow forecasting with global climate change implications—A comparative study between genetic programming and neural network models. *Journal of Hydrology*, 352: 336-354.

Mattera, D. S. Haykin. 1999. Support vector machines for dynamic reconstruction of a chaotic system. *In*: S. Bernhard, I. Ikonf, J.C.B. Christopher and J.S. Alexander (eds). *Advances in kernel methods*. MIT Press: 211-241

Nayak, P. C., K. Sudheer, D. Rangan and K. Ramasastri. 2004. A neuro-fuzzy computing technique for modeling hydrological time series. *Journal of Hydrology*, 291: 52-66.

Pierce, M. W., C. I. Thornton, and S. R. Abt. 2009. Predicting peak outflow from breached embankment dams. *Journal of Hydrologic Engineering*, 15: 338-349.

Ponce, V. M., A. Taher-Shamsi and A. V. Shetty. 2003. Dam-breach flood wave propagation using dimensionless parameters. *Journal of Hydraulic Engineering*, 129: 777-782.

Ranjithan, S., J. Eheart and J. Garrett. 1993. Neural network-based screening for groundwater reclamation under uncertainty. *Water Resources Research*, 29: 563-574.

Sattar, A. M. and B. Gharabaghi. 2015. Gene expression models for prediction of longitudinal dispersion coefficient in streams. *Journal of Hydrology*, 524: 587-596.

Schölkopf, B., C. Burges and V. Vapnik. 1995. Extracting support data for a given task. *In*: Fayyad, U. M. and Uthurusamy, R. (eds.) Proceedings, First International Conference on Knowledge Discovery & Data Mining. Menlo Park, CA: AAAI Press.

Service, S. C. 1981. Simplified Dam-breach Routing Procedure, USDA, Washington DC, Technical Release No. 66 (Rev.1).

Singh, V. P. and P. D. Scarlatos. 1988. Analysis of gradual earth-dam failure. Journal of hydraulic engineering, 114: 21-42.

Sivapragasam, C., R. Maheswaran and V. Venkatesh. 2008. Genetic programming approach for flood routing in natural channels. Hydrological processes, 22: 623-628.

Smith, J. and R. N. Eli. 1995. Neural-network models of rainfall-runoff process. Journal of water resources planning and management, 121: 499-508.

Smola, A. J. 1996. Regression estimation with support vector learning machines. Master's thesis, Technische Universit at Munchen.

Soil Conservation Service .1981. Simplified Dam-breach Routing Procedure. Technical Release No. 66 (Rev.1). USDA, Washington, DC, p. 39.

Tokar, A. S. and P. A. Johnson. 1999. Rainfall-runoff modeling using artificial neural networks. Journal of Hydrologic Engineering, 4: 232-239.

Tsai, C. W. 2005. Flood routing in mild-sloped rivers—wave characteristics and downstream backwater effect. Journal of Hydrology, 308, 151-167.

U.S. Bureau of Reclamation .1982. Guidelines for Defining Inundated Areas downstream from Bureau of Reclamation Dams, Reclamation Planning Instruction No. 82-11, June 15, 1982.

Vaskinn, K. A., A. Lovoll, K. Hoeg, M. Morri, G. Hanson and M. Hassan. 2004. Physical modeling of breach formation: Large scale field tests. Proceedings of the Dam safety.

Vapnik, V. 1995. The Nature of Statistical Learning Theory. Springer.

Wahl, T. 1998. Prediction of embankment dam breach parameters—a literature review and needs assessment. Dam Safety Rep No. DSO-98-004, US Dept. of the Interior, Bur of Reclamation, Denver, CO.

Wang, W. C., K. W. Chau, C. T. Cheng and L. Qiu. 2009. A comparison of performance of several artificial intelligence methods for forecasting monthly discharge time series. Journal of hydrology, 374: 294-306.

Whigham, P. and P. Crapper. 2001. Modelling rainfall-runoff using genetic programming. Mathematical and Computer Modelling, 33: 707-721.

Xia, J., B. Lin, R. A. Falconer and G. Wang. 2010. Modelling dam-break flows over mobile beds using a 2D coupled approach. Advances in Water Resources, 33: 171-183.

Xu, Y. and L. Zhang. 2009. Breaching parameters for earth and rockfill dams. Journal of Geotechnical and Geoenvironmental Engineering, 135: 1957-1970.

Yu, P. S., S. T. Chen and I. F. Chang. 2006. Support vector regression for real-time flood stage forecasting. Journal of Hydrology, 328: 704-716.

Evaluation of Predictive Potential of a learning Machine Method on Peak Outflow Due To Embankment Dam Failure

Hamed Farhadi^۱, Kazem Esmaili^۲

Abstract

In this study, the potential of Support Vector Machine (SVM) in the prediction of peak outflow due to dam failure was evaluated. Huge volume deposited behind dam may cause a large casualty in case of sudden release of it to downstream. Head and volume of water at the time of failure were considered as inputs to SVM model. To train these models, 70% of 112 gathered data from literature was used as training subset and the rest 30% was used to test the model as test subset, at the same time these two subsets were selected in a way to be statistically similar. After studying four SVM models with different kernel functions, i.e. Linear, Polynomial, Radial Basis Function and Sigmoid, it was found that SVM with Radial Basis Kernel function outperforms other models. Results of the statistical evaluation of the proposed model are satisfying with $R^2=0.96$, RMSE=0.03 for training phase and $R^2=0.94$, RMSE=0.05 for the test phase. A comparison was made between some conventional empirical equations and the proposed model in which results shows proposed SVM model surpasses empirical equations in predicting peak outflow due to embankment dam failure.

Keywords: Learning Machine, Artificial Intelligence, Dam break, Earth dam, Flood peak outflow

1- PhD student of hydraulic structures, water science and engineering department, Ferdowsi University Of Mashhad, farhadi.edu@gmail.com

۲- Associate Professor, water science and engineering department, Ferdowsi University of Mashhad, esmaili@um.ac.ir (corresponding author)