بسم الله الرحمن الرحيم

دانشگاه صنعتی اصفهان

دانشکده مهندسی عمران

کاربرد ماشین‌های بردار پشتیبان در مدلسازی منابع آب

پایان نامه کارشناسی ارشد مهندسی آب

محسن بهزاد جزی

استاد راهنما

دکتر کیوان اصغری

۱۳۸۶

دانشگاه صنعتی اصفهان

دانشکده مهندسی عمران

پایان نامه کارشناسی ارشد عمران-گرایش مهندسی آب
آقای محسن بهزاد جزی

تحت عنوان

کاربرد ماشین‌های بردار پشتیبان در مدلسازی منابع آب

در تاریخ ۱۳۸۶/۱۲/۱۱ توسط کمیته تخصصی زیر مورد بررسی و تصویب نهایی قرار گرفت.

۱- استاد راهنمای پایان نامه	دکتر کیوان اصغری
۲- استاد مشاور پایان نامه	دکتر مازیار پالهنگ
۳- استاد داور	دکتر حمیدرضا صفوی
۴- استاد داور	دکتر سعید اسلامیان
۵- سرپرست تحصیلات تکمیلی دانشکده	دکتر مرتضی مدح‌خوان

**کلیه حقوق مادی مترتب بر نتایج مطالعات،
ابتکارات و نوآوریهای ناشی از تحقیق موضوع
این پایان نامه (رساله) متعلق به دانشگاه صنعتی
اصفهان است.**

فهرست مطالب

عنوان **صفحه**

فهرست مطالب شش

چکیده ۱

فصل اول: مقدمه

۱-۱ معرفی موضوع ۲

۱-۲ مدلهای هیدرولوژیکی ۳

۱-۳ اهداف تحقیق ۵

فصل دوم: مروری بر تحقیقات انجام شده

۲-۱ مقدمه ۷

۲-۱-۱ شبکه‌های عصبی مصنوعی (ANNs) ۷

۲-۱-۲ ماشین‌های بردار پشتیبان (SVMs) ۸

۲-۲ برتری روش SVMs نسبت به ANNs ۸

۲-۳ ماشین‌های بردار پشتیبان و مدلسازی منابع آب ۱۱

فصل سوم: ماشین‌های بردار پشتیبان

۳-۱ مقدمه ۲۰

۳-۲ کمینه‌سازی خطا ۲۱

۳-۲-۱ اصل استقرایی کمینه‌سازی خطای تجربی (ERM) ۲۱

۳-۲-۲ اصل استقرایی کمینه‌سازی خطای ساختاری (SRM) ۲۲

۳-۲-۳ هموارسازی ۲٤

۳-۳ ماشین‌های بردار پشتیبان ۲۵

۳-۳-۱ شکل رگرسیونی ماشین‌های بردار پشتیبان ۲۵

۳-۳-۲ تابع هدف مبنا ۲۷
۳-۳-۳ نمایش دوگانه تابع هدف مبنا ۲۷
۳-۴ یادگیری توابع غیر خطی ۳۰
۳-۴-۱ یادگیری در فضای ویژگی ۳۱
۳-۴-۲ نگاشت ضمنی به فضای ویژگی ۳۳
۳-۴-۳ تئوری Mercer ۳٤
۳-۴-۴ تصویر کلی از ساختار SVR ۳۶
۳-۵ به‌کارگیری ماشین‌های بردار پشتیبان ۳۶
۳-۵-۱ تعیین تابع کرنل مناسب ۳۷
۳-۵-۲ تعیین وراپارامترهای ماشین بردار پشتیبان ۳۸
۳-۵-۳ نرمالسازی داده‌ها ۴۰
۳-۵-۴ حل مسئله‌ی بهینه سازی درجه دوم مقید محدب ۴۰

فصل چهارم: کاربرد ماشین‌های بردار پشتیبان در مدلسازی منابع آب

۴-۱ مقدمه ۴۲
۴-۲ پیش‌بینی کوتاه‌مدت رواناب ۴۲
۴-۲-۱ معرفی منطقه‌ی مورد مطالعه ۴۳
۴-۲-۲ هدف و روش تحقیق ۴۶
۴-۲-۳ نتایج ۴۷
۴-۳ پیش بینی ارتفاع سطح آب در چاه مشاهده‌ای ۵۲
۴-۳-۱ معرفی منطقه‌ی مورد مطالعه ۵۴
۴-۳-۲ هدف و روش تحقیق ۵۶
۴-۳-۳ نتایج ۵۸
۴-۴ پیش بینی زمانی- مکانی بارندگی ۶۴
۴-۴-۱ معرفی منطقه‌ی مورد مطالعه ۶۵
۴-۴-۲ هدف و روش تحقیق ۶۷
۴-۴-۳ نتایج ۶۸

فصل پنجم: خلاصه و پیشنهادات

۵-۱ خلاصه ۷۰

۵-۲ پیشنهادات ۷۲

پیوست

پیوست اول: اندازه‌گیری خطای تخمین ۷۴

پیوست دوم: ارزیابی عملکرد مدلها ۷۸

مراجع **۸۲**

چکیده

کمبود آب، تغییرات آب و هوایی و عدم قطعیت‌های هیدرولوژیکی ضرورت وجود مدلسازی و مدیریت هدفمند منابع آب را آشکار می‌سازند. مهار کردن و استفاده صحیح از منابع آبهای جاری، توسعه منابع آب زیرزمینی، کاهش اثرات منفی ناشی از سیل و یا خشکسالی و تأمین آب آشامیدنی سالم نیازمند مدلهایی با قابلیت انجام پیش‌بینی‌های دقیق ومطمئن هستند. مدلسازی داده–محور، از جمله روشهای نوینی است که به سرعت در حال گسترش در زمینه‌های متنوع علمی می‌باشد. این شیوه‌ی مدلسازی می‌تواند در مسائل گوناگونی جایگزین سایر روشهای شبیه‌سازی نظیر مدلسازی فیزیکی ویا مدلسازی تجربی شود.

این مطالعه، اصول روشی جدید و پیشرفته به نام ماشین‌های بردار پشتیبان را که بر پایه تئوری یادگیری آماری استوار است، مورد بحث قرار خواهد داد. استفاده از این روش یادگیری باعث افزایش قابلیت عمومیت‌پذیری ماشین خواهد شد که منجر به بهبود یافتن دقت مدل در مقایسه با سایر روشهای داده–محور پیشین شده است.

هدف این پژوهش آشنایی با مفهوم ماشین‌های بردار پشتیبان، به منظور دستیابی به فرایندهای فیزیکی پیچیده و رفتارهای غیر خطی سیستم‌های هیدرولوژیکی می‌باشد. در این پایان نامه عملکرد روش یادگیری مذکور در قالب سه کاربرد مختلف از مدلسازی منابع آب شامل ۱) پیش‌بینی کوتاه مدت رواناب، ۲) تخمین سطح تراز آب در چاه مشاهده‌ای و ۳) پیش‌بینی زمانی–مکانی بارندگی، مورد بررسی قرار گرفته است. بدین جهت ترکیبهای مختلفی از داده‌ها برای پیش‌بینی رفتار سیستم‌های هیدرولوژیکی گفته شده، معرفی می‌گردند. مقایسه نتایج روش پیشنهادی با نتایج مدل شبکه‌های عصبی مصنوعی جهت ارزیابی کارایی آن، قابلیت پیش‌بینی بالای ماشین‌های بردار پشتیبان را در کاربردهای فوق، روشن می‌سازد. به عنوان نمونه استفاده از این نوع ماشین یادگیر در تخمین رواناب موجب شده تا جذر مربع میانگین خطا در دو مدل ساخته شده به مقدار ۱۶ و ۲۷ درصد نسبت به روش شبکه‌های عصبی مصنوعی کاهش یابد. در پنج مدل ارائه شده در کاربرد دوم و همچنین سه مدل پیش‌بینی بارندگی در کاربرد سوم نیز این کاهش خطا مقدار متوسطی به ترتیب برابر با ۱۲ و۷۲ درصد داشته است. عملکرد موفقیت‌آمیز این نوع ماشین یادگیر در این تحقیق، امکان توسعه و استفاده از آن را در کاربردهای دیگری از مدلسازی منابع آب نشان می‌دهد.

فصل اول
مقدمه

۱-۱ معرفی موضوع

آب در همه زمانها به عنوان پشتوانه‌ای در دستیابی بـه اهـداف اجتمـاعی، مهـار شـده و مـورد اسـتفاده قـرار گرفته است. با این وجود واضح است که تلاشهای انجام شده در جهت استفاده پایـدار از منـابع آب ناکـافی و یا بی‌هدف بوده است. به علاوه در آینده به علت تقاضای روزافزون آب، منـابع آب موجـود بـا فشـار بیشـتری مواجه خواهند بود. بنابراین برای رسیدن به اهداف توسعه، مدیریت هدفمند منابع آب بیش از پـیش مـورد نیـاز است.

ایران سرزمینی خشک با نزولات جوی بسیار کم مـی‌باشـد بـه طوریکـه اگـر میـانگین بارنـدگی سـالانه در سطح کره زمین را که حدود ۸۶۰ میلی‌متر تخمین زده مـی‌شـود بـا متوسـط بارنـدگی سـالانه ایـران کـه تقریبـاً رقمی معادل ۲۵۲ میلی‌متر است مقایسه کنیم، ملاحظه خواهد شد کـه بارنـدگی در ایـران حتـی کمتـر از یـک سوم متوسط بارندگی در سطح دنیاست. علاوه بر این زمان ریزش نزولات جوی و محل ریزش آنها نیز بـا نیـاز بخش کشاورزی، که مصرف کننده اصلی آب در کشور می‌باشد، مطابقت نـدارد. بنـابراین بایـد پـذیرفت کـه خشکی در ایران یک واقعیت اقلیمی است که باعث گردیده با در نظر گرفتن جمعیت کنونی کشـور و مقـدار سرانه آب تجدید شونده‌ی حدود ۱۵۰۰ متر مکعب در سال، جزو کشورهای با تنش آب تلقـی شـود. بـا رونـد کنونی رشد جمعیت و مصرف آب پیش‌بینی مـی‌شـود در سـال ۱۴۰۰ سـرانه آب تجدیـد شـونده‌ی کشـور بـه کمتر از ۱۲۵۰ متر مکعب در سال برسد که در آن صورت کشور در ردیف کشورهای کـم آب قـرار خواهـد گرفت و با مشکلات زیادی مواجه خواهیم بود.

یکی از راه‌های سازگاری با کم‌آبی استفاده بهینه از منابع آب و افزایش بهره‌وری آب است. باید سعی نمود تا حد ممکن از نزولات جوی، جریان آبهای سطحی، منابع آب زیرزمینی و رطوبت خاک به نحو مطلوب و بهینه استفاده شود و اینکار عملی نخواهد بود مگر با ایجاد مدیریت هدفمند، مطمئن و جامع منابع آب [۱].

این چنین مدیریتی به مدلهایی با قابلیت پیش‌بینی مطمئن، همراه با سیستم‌های تصمیم‌گیرنده نیازمند است. از طرفی پیشرفتهای سریع در زمینه‌های فناوری اطلاعات، مفاهیم ترکیب داده‌ها و الگوریتم‌های یادگیری (تئوری یادگیری محاسباتی و مدلسازی داده–محور) تغییرات اساسی در مدیریت منابع آب بوجود آورده‌اند. در ادامه مروری بر انواع مدلهای به کارگرفته شده در مدلسازی منابع آب خواهیم داشت.

۱-۲ مدلهای هیدرولوژیکی

هیدرولوژی مطالعه علمی آب و خصوصیات آن، توزیع و تأثیرات آن بر روی سطح زمین، خاک و جو می‌باشد. اکثر فرایندهای هیدرولوژیکی به شدت غیرخطی و شامل تغییرات زمانی و مکانی می‌باشند. علاوه بر این عدم قطعیت پارامترهای هیدرولوژیکی به میزان پیچیدگی آنها می‌افزاید. معمولاً هیدرولوژیست‌ها با مشکلاتی مربوط به پیش‌بینی و تخمین کمیت‌هایی نظیر رواناب، بارندگی، ارتفاع سطح آب، غلظت مواد آلاینده و ... مواجه هستند. اینگونه اطلاعات هیدرولوژیکی در طراحی‌های مربوط به مدیریت منابع آب مورد نیاز می‌باشند.

مدلهای هیدرولوژیکی که امروزه استفاده می‌گردند به سه دسته تقسیم‌بندی می‌شوند: مدلهای تجربی، ژئومورفولوژیکی و فیزیکی [۲]. مدلهای تجربی که سیستم هیدرولوژیکی (به مانند حوضه‌ی آبریز) را همانند یک دستگاه پیچیده‌ای که قادر به فهم فرایندهای داخل آن نبوده[1] در نظر می‌گیرند و سعی می‌شود تا با استفاده از مدل، رابطه ای بین متغیرهای ورودی (مثل بارندگی، درجه حرارت و ...) و خروجی (رواناب حوضه‌ی آبریز) استنباط شود. اینگونه مدلها بر مبنای مفهوم فیزیکی خاصی پایه ریزی نشده‌اند و به داده‌های فراوانی از رفتار سیستم در درازمدت نیاز دارند. لذا در حوضه هایی که ایستگاههای اندازه‌گیری موجود نیست، کاربردی ندارند. مدلهای ژئومورفولوژیکی مدلهای پیشرفته‌تری نسبت به مدلهای تجربی هستند که نسبتاً ساختار حوضه‌های آبریز و شبکه‌ی جریان سطحی را به خوبی بیان می‌کنند، اما به علت استفاده از فرضهای ساده کننده (به مانند پاسخ خطی اجزای حوضه‌ی آبریز) در مورد سیستم، از دقت کافی برخوردار نمی‌باشند. مدلهای فیزیکی تا آنجا که قادر باشند می‌توانند فرایندهای فیزیکی و سایر فرایندهای موجود در سیستم را در غالب معادلات دیفرانسیلی جزئی نشان دهند که تحلیل آنها نیازمند گسسته‌سازی زمانی و مکانی دامنه‌ی مسئله و بهره‌گیری از روش‌های حل عددی خواهد بود. این موضوع دلالت بر این واقعیت دارد که

[1] Black Box

مدلهای فیزیکی زمانی مفید واقع می‌گردند که داده‌های مربوط به خصوصیات فیزیکی سیستم آبی در نقاط دامنه‌ی مسئله موجود باشند. اما این گونه داده‌ها حتی در سیستم‌های آبی تحقیقاتی (با وسایل اندازگیری مجهز) نیز به ندرت یافت می شوند. با این وجود رفتار فیزیکی سیستم آبی در مواردی به طور واضح (به دلیل اغتشاشات ، غیر خطی بودن و یا غیر پایدار بودن فرایندها) قابل فهم نبوده و یا مدلها آنچنان ساده شده می‌باشند که استفاده از آنها جهت تحلیل سیستم به پاسخ‌های نادرست منجر خواهد شد. سروشیان[۲] (۱۹۹۴) مروری بر مدلهای ساده و پیچیده‌ی فیزیکی برای پیش‌بینی رواناب حوضه‌ی آبریز داشته است. در این بررسی، اگرچه بدون واسنجی مدل پیچیده، تخمین رواناب دقت بیشتری نسبت به مدل ساده داشته اما پس از واسنجی دقت هر دو مدل یکسان بوده است. او بر طبق این نتیجه دریافت که پیچیدگی مدل فیزیکی نمی‌تواند دلالت بر دقت آن داشته باشد. او همچنین عدم وجود داده‌های مکانی کافی مربوط به بارندگی را دلیل عمده کاهش دقت شبیه سازی دانسته و بر ضرورت استفاده از واسنجی در مدلسازی های فیزیکی تاکید نموده است [۳]. با توجه به این مطالعه، به طور کلی دو مشکل در ارتباط با استفاده از مدلهای فیزیکی وجود دارد: (۱) در عمل تقریباً همیشه پارامترهای مدل باید کالیبره گردند، در حالی که در اکثر موارد، داده‌های کافی برای کالیبراسیون وجود ندارد و (۲) فرایندهای فیزیکی و دیگر فرایندهای موجود در سیستم به طور کامل منظور نمی‌شوند. علیرغم مسائل و مشکلات ذکر شده، مدلهای فیزیکی یا مفهومی در درک فرایندهای هیدرولوژیکی پراهمیت هستند.

در گذشته، محققین هنگامی که با تعداد محدودی از داده‌ها مواجه بوده و یا درک درستی از فرایندهای موجود در سیستم وجود نداشت، از مدلهای تجربی استفاده می‌کردند. از معمول‌ترین مدلهای تجربی موجود مدلهای رگرسیونی را می‌توان نام برد. این مدلها الگوی ریاضی خاصی برای سیستم در نظر گرفته و سپس پارامترهای مجهول بر مبنای کمینه‌سازی اختلاف موجود بین مقادیر پیش‌بینی شده و مقادیر اندازه‌گیری شده (معمولاً به روش حداقل مربعات) محاسبه می‌شوند. انجمن هوش مصنوعی[۳] با مشاهده ضعف ناشی از در نظر گرفتن یک رابطه‌ی ریاضی پیش از حل مسئله در این مدلها، شبکه‌های عصبی مصنوعی[۴] (ANNs) را به عنوان روش دیگری برای مدلسازی پیشنهاد نمود. این روش مشابه با مدلهای رگرسیونی است؛ بنابراین می توان آنها را در دسته مدلهای تجربی قرار داد اما در روند مدلسازی شکل خاصی از توابع ریاضی را در نظر نمی‌گیرد. ANNs رفتار یک سلول عصبی زنده را در شکل ساده‌ی آن تقلید می‌کند. به علاوه به دلیل اینکه فرد در انتخاب تعداد لایه‌های مخفی و یا تعداد نرونهای شبکه عصبی (که ساختار شبکه را معین می کنند) آزاد است، یک وسیله چند کاربردی محسوب می‌گردد. این چنین ویژگیهای خاص سبب افزایش کاربردهای ANNs در منابع آب در دو دهه قبل شده است [۴]. ANNs راه‌حل قدرتمندی در حل مسائل

[2] Sorooshian
[3] Artificial Intelligence Community
[4] Artificial Neural Networks

محسوب می‌شود اما محدودیتهایی نیز دارد. از جمله این محدودیتها (۱) عدم وجود یک توجیه فیزیکی برای ساختار مدل، (۲) نبود روش ثابتی برای تعیین ساختار شبکه‌ی عصبی، (۳) احتمال همگرا شدن جواب بهینه در یک کمینه‌ی محلی و (۴) وابستگی صحت نتایج حاصل از مدلسازی به کمک شبکه‌های عصبی به کیفیت وکمیت داده‌های پیشین را می‌توان نام برد [۲]. برخی از این ضعف‌ها با استفاده از الگوریتم‌های جدید یادگیری از جمله ماشین‌های بردار پشتیبان (که موضوع این مطالعه است) برطرف شده‌اند. برای روشن‌تر شدن این موضوع، در فصل دوم به بررسی جامع‌تری از مزایای روش پیشنهادی در مقایسه با ANNs خواهیم پرداخت.

۱-۳ اهداف تحقیق

مدیریت موفق منابع آب به روشهای جهت‌دار، جامع و سیستماتیک نیازمند است تا بدینگونه اطلاعات مورد نیاز برای تصمیم‌گیری با کمترین هزینه در اختیار مدیران قرارگیرد. مدلهایی که قادر به یادگیری و تعمیم مفاهیم هستند می‌توانند در مسائل مربوط به تخمین، پیش‌بینی، مدیریت و کنترل، در جنبه‌های مختلفی از منابع آب مورد استفاده قرار گیرند. این نوع مدلها در مقابل مدلهای فیزیکی و یا مدلهای رفتار-محور[5] فراتر از درک محدود ما از رفتار یک سیستم و با استفاده از تشخیص الگوهای مشاهده شده در زوج‌های ورودی و خروجی، اطلاعات لازم را استخراج می‌کنند. این قابلیت تشخیص الگوهای مختلف، می‌تواند در قالب مدلهایی که توانایی تقلید رفتارهای واقعی حاکم بر فرایندهای فیزیکی سیستم را دارند، توسعه یابد.

به طور عمده، هدف این پایان‌نامه بررسی و ارزیابی امکان استفاده از یکی از جدیدترین الگوریتم‌های یادگیری به نام ماشین‌های بردار پشتیبان[6] (SVMs)، به عنوان مکمل و یا جانشین روش‌های پیشین در حل مسائل پیچیده‌ی منابع آب می‌باشد. در این راستا اهداف اصلی این مطالعه شامل موارد زیر خواهند بود:

۱ -بررسی و شناسایی قابلیت‌های ماشین‌های بردار پشتیبان،

۲ -توسعه مدلهایی بر مبنای روش پیشنهادی، به منظور شبیه‌سازی‌های هیدرولوژیکی نظیر:

- پیش‌بینی کوتاه‌مدت رواناب
- تخمین تراز آب زیرزمینی در چاه مشاهده‌ای
- پیش‌بینی زمانی- مکانی بارندگی

۳ -بررسی عملکرد روش مذکور و مقایسه‌ی نتایج حاصل از آن با نتایج موجود از روش شبکه‌های عصبی مصنوعی.

---

[5] Behavior-Driven Models
[6] Support Vector Machines

در فصل بعدی این پژوهش، مروری بر تحقیقات پیشین در زمینه‌ی کاربردهای ماشین‌های بردار پشتیبان در مدلسازی‌های هیدرولوژیکی و منابع آب صورت پذیرفته است. در فصل سوم به کلیاتی پیرامون الگوریتم یادگیری ماشین‌های بردار پشتیبان و نحوه به‌کارگیری آن پرداخته می‌شود. در فصل چهارم مدلسازی سه کاربرد هیدرولوژیکی متفاوت توسط روش پیشنهادی ارائه شده و نتایج آن مورد بررسی قرار گرفته است و نهایتاً در آخرین فصل این پایان نامه خلاصه‌ی نتایج و پیشنهاداتی جهت انجام پژوهش‌های آتی، بیان می گردد.

فصل دوم
مروری بر تحقیقات انجام شده

۲-۱ مقدمه

به طور کلی الگوریتم‌های یادگیری آماری[۷] که SVMs نوع خاصی از این الگوریتم‌ها است از جمله روش‌های یادگیری با نظارت[۸] محسوب می‌شوند که یادگیری را بر اساس دسته‌ای از داده‌های آموزشی (زوج‌های ورودی و خروجی) انجام می‌دهند. مطالبی که در مورد ماشین‌های بردار پشتیبان موجود است شامل مقالات متنوعی در ارتباط با به‌کار گیری، بهبود عملکرد و توسعه مفهوم آن می‌باشند و بیانگر توانایی برتر این ابزار در تسریع روند مدلسازی و ویژگی عمومیت‌پذیری هستند. در ادامه به بررسی دو نمونه از الگوریتم‌های یادگیری مورد استفاده در مدلسازی منابع آب شامل شبکه‌های عصبی مصنوعی و ماشین‌های بردار پشتیبان خواهیم پرداخت.

۲-۱-۱ شبکه‌های عصبی مصنوعی (ANNs)

تحقیقات فراوانی صورت پذیرفته تا شبکه‌های عصبی مصنوعی را جانشین مدلهای ریاضی نماید. ANNs به عنوان یک الگوریتم موفق به طور گسترده‌ای در هیدرولوژی استفاده می‌شود. روابط پیچیده غیرخطی موجود در فرایندهای هیدرولوژیکی که به عوامل متعددی وابسته است توسط این روش مورد بررسی قرار گرفته و کارایی قابل قبولی را از خود به نمایش گذاشته است [۴ و ۲].

---

[7] Statistical Learning Algorithms
[8] Supervised Learning

۲-۱-۲ ماشین‌های بردار پشتیبان (SVMs)

ماشین‌های بردار پشتیبان ابزاری است که بر اساس اصول تئوری یادگیری آماری[۹] توسط وَپنیک[۱۰] (۱۹۸۲، ۱۹۹۲، ۱۹۹۵، ۱۹۹۸) توسعه یافت [۵]. هنگامی که SVMs برای اولین بار معرفی گردید، از استقبالی مشابه آنچه در ابتدا متوجه ANNs بود، روبرو شد و به عنوان زمینه‌ی فعالی در تحقیقات یادگیری ماشین مدنظر محققین قرار گرفت. SVMs بر اساس مفاهیم کنترل ظرفیت توابع[۱۱] و یا اصل کمینه سازی خطای ساختاری[۱۲] (SRM) بنیانگذاری شده است. در این روش، جستجوی انجام شده در روند مدلسازی برای یافتن تابع هدف (تابع تخمین ویا تابع تصمیم)، به گونه‌ای است که همزمان با بیشینه نمودن کارایی و خاصیت تعمیم پذیری مدل میزان پیچیدگی آن کمینه گردد. این عمل با بهره‌گیری از داده‌های آموزشی ورودی و خروجی صورت می‌پذیرد.

۲-۲ برتری روش SVMs نسبت به ANNs

هما نطور که پیشتر گفته شد در دهه نود ANNs به طور گسترده‌ای در کابردهای هیدرولوژیکی متنوعی استفاده شده است. در مروری بر این کاربردها که توسط هیئت بررسی ASCE[۱۳] در سال ۲۰۰۰ انجام گرفته به این نکته اشاره شده که با وجود مزایایی که ANNs دارا است، از محدودیت‌های عمده‌ای نیز برخوردار است که باعث ایجاد برخوردهای شکاکانه‌ای نسبت به این روش تحقیق شده است. اما به نظر می‌رسد SVMs می‌تواند جایگزین مناسبی برای غلبه بر برخی از محدودیتهای شبکه‌ی عصبی باشد، در حالی که نقاط قوت ANNs پا برجا باقی بماند. در ادامه مزایای SVMs با توجه به مسائلی که توسط ASCE و همچنین لیونگ و همکارش (۲۰۰۲) [۶] به عنوان کاستی‌های ANNs مطرح شده‌اند، ارائه می‌گردد.

- ANNs، مدل جعبه سیاه

در ANNs مجموعه‌ی مقادیر وزنها و سایر پارامترها (بعد از آموزش) هیچگونه اطلاعاتی را برای کاربر آشکار نمی‌سازند. این مسئله مهمترین دلیل برای ایجاد دیدی شک برانگیز به این روش شده است. در حالی که SVMs برخلاف ANNs مدل جعبه سیاه نیست. این مسئله می‌تواند به صورت تئوریک با استفاده از مفاهیم تئوری یادگیری محاسباتی تحلیل گردد. مقادیر نهایی ضرایب لاگرانژ مشخص کننده اهمیت نسبی داده‌های آموزش در دستیابی به تابع تخمین می‌باشد. به طور کلی SVMs توصیفی کلی در مورد رسیدن به تابع تخمین عرضه می‌کند.

---

[9] Statistical Learning Theory
[10] Vapnik
[11] Capacity Control
[12] Structural Risk Minimization
[13] American Socity of Civil Engineering

- تشخیص مجموعه‌ی داده‌های آموزش بهینه

همانطور که توسط ASCE ذکر شده، عوامل بازدارنده‌ای ناشی از دقت و هزینه‌ی مربوط به جمع‌آوری داده‌ها در بسیاری از کاربردهای هیدرولوژیکی وجود دارد. با توجه به اینکه ANNs نیازمند تعداد زیادی از داده است، لذا بدون وجود داده‌هایی که از لحاظ کمی و کیفی مناسب مدلسازی باشند، ویژگی عمومیت‌پذیری ANNs دارای عملکرد ضعیفی خواهد بود. از آنجایی که SVMs بر اصل کمینه‌سازی خطای ساختاری (SRM) استوار است، بنابراین از لحاظ ویژگی عمومیت‌پذیری عملکرد بهتری در مقایسه با ANNs نشان می‌دهد. به علاوه همانطور که در ادامه خواهیم دید، به دلیل استفاده SVMs از تابع خطای خاصی، تعداد الگوهای آموزش که برای تعریف تابع تخمین استفاده می شوند، بخش کوچکی از کل داده‌ها خواهند بود و لذا این امکان فراهم می‌گردد که تنها این بخش از داده‌ها (و نه کلیه‌ی داده‌های آموزش) ذخیره شود .

- بهبود تحلیل سری‌هایی زمانی

یکی از مهمترین مسائلی که در تحلیل سری‌های زمانی مورد توجه قرار می‌گیرد تاخیرهای زمانی مورد نیاز است. بدین معنا که در بسیاری از مطالعات مربوط به آنها می‌بایست از داده‌هایی با تاخیرهای زمانی متعدد استفاده نمود، اگر چه در وحله‌ی اول روشن نیست که تا چه مدت زمانی نیاز به بازگشت به گذشته است تا تأثیرات زمانی داده‌ها در نظر گرفته شود. به هر حال این مسئله باعث افزایش پارامترهای ANNs و پیچیدگی ساختار آن خواهد شد. اما SVMs می‌تواند با افزایش تعداد خصیصه‌ها به راحتی برخورد کند چرا که به دلیل بهره‌گیری از ضرب داخلی بردارها در روند مدلسازی، به کارگیری بردارهای ورودی با هر تعداد بعدی به آسانی امکان پذیر می‌باشد.

- ساختار بهینه

SVMs برخلاف ANNs ساختار نهایی خود را به صورت خودکار و با حل مسئله بهینه‌سازی مربوطه استخراج می‌نماید. در حالی که انتخاب ساختار ANNs نیاز به روند سعی و خطای وقت‌گیر و نیز تجربه‌ی بالای کاربر دارد. به عبارت دیگر می‌توان تعداد بردارهای پشتیبان را که در تعریف تابع هدف استفاده می‌شود، معادل با تعداد نرونهای لایه‌ی مخفی دانست با این تفاوت که این بردارها به صورت خودکار حاصل می‌شوند.

- کمینه‌ی جهانی به جای کمینه‌ی محلی

مسئله‌ی بهینه‌سازی که در راستای ایجاد مدل SVM شکل می‌گیرد دارای جواب یکتایی است و لذا این جواب جهانی خواهد بود. در حالی که ممکن است روند حل مدل ANN به کمینه‌های محلی منجر شود.

- استفاده از بردار ورودی با هر تعداد خصیصه

در انتخاب بردار ورودی، به طور معمول سعی می‌شود تا کوچکترین مجموعه‌ای از خصیصه‌ها که اطلاعات ضروری را بهمراه دارد انتخاب شود. که این عمل کاهش بعد نامیده می‌شود. با توجه به اینکه

افزایش تعداد خصیصه‌ها می‌تواند منجر به افزایش پیچیدگی مدل، کاهش عملکرد عمومیت‌پذیری آن و بالابردن هزینه محاسباتی شود، استفاده از این روند مفید به نظر می‌رسد. به عنوان نمونه با استفاده از آنالیز مولفه اصلی[۱۴] (PCA)، کاهش بعد از طریق خارج ساختن خصیصه‌های مربوط به جهاتی که داده‌ها تغییرات کمتری دارند انجام می‌شود. اگر چه هیچگونه تضمینی برای عدم ضرورت آن خصیصه در تحلیل سیستم وجود ندارد. SVMs توانسته این مشکل را با ارائه راه حلی کارآمد، از طریق شکل دوگانه خود (در قالب نمایش داده‌ها به صورت ضرب داخلی) و به کارگیری توابع کرنل برطرف سازد. بنابراین می‌توان با استفاده از توابع کرنل با افزایش تعداد خصیصه‌ها (بعد بردار ورودی) برخورد نمود.

- تاثیر مقادیر اولیه وزنها بر نتایج شبکه عصبی

یکی از مواردی که در تعیین ساختار ANN اهمیت دارد، مقدار دهی اولیه به وزنهای مرتبط با نرونهای موجود در شبکه عصبی است. با توجه به اینکه در بیشتر اوقات به منظور جلوگیری از ورآموزی، آموزش شبکه عصبی در زمان معینی متوقف خواهد شد لذا مقادیر متفاوت وزنهای اولیه می‌تواند نتایج کم و بیش متفاوتی ایجاد نماید که در بسیاری از موارد این موضوع نادیده گرفته می‌شود. با این وجود برخی از محققین با آموزش شبکه با مقادیر متفاوت وزنهای اولیه و سپس یافتن بهترین شبکه بر اساس نتایج مرحله آزمون و یا میانگین‌گیری از نتایج موجود، تا حدودی بر این مشکل غلبه می نمایند. به هر جهت در معرفی شبکه توسعه یافته توسط محقق ذکر مقادیر اولیه‌ی وزنها ضروری به نظر می رسد. این در حالی است که با دانستن مجموعه‌های آموزش و آزمایش و همچنین مقادیر پارامترهای بهینه SVM بدون توجه به نقاط اولیه، مدل تنها به یک جواب جهانی همگرا خواهد شد.

- حذف اغتشاشات از داده‌ها توسط SVM

همانطور که در ادامه خواهیم دید، استفاده از تابع خطای ε-*insensitive* در روند توسعه مدل SVM باعث فیلتر نمودن اغتشاشات (noises) می‌گردد. در صورتی که اگر شبکه‌های عصبی به طور مناسب آموزش داده نشود، یادگیری اینگونه اغتشاشات منجر به ایجاد ورآموزی شبکه می‌شود.

ذکر این نکته ضروری است که در مدلهای داده-محور زمان قابل توجهی نیاز است تا مدل به ساختار مناسبی دست یابد. اگر چه هنگامی که مدل آموزش داده شد، استفاده از مدل به منظور پیش‌بینی‌های آینده در کسر کوچکی از زمان لازم برای آموزش صورت می‌پذیرد. البته باید خاطر نشان شد که زمان لازم برای یافتن جواب بهینه به تعداد داده‌های مسئله و بعد آن (تعداد خصیصه های بردار ورودی) وابسته است. به طور کلی ممکن است یافتن جواب بهینه با افزایش تعداد داده‌ها و/یا افزایش بعد آنها دچار مشکل شود.

[14] Principle Componenet Analysis

۲-۳ ماشین‌های بردار پشتیبان و مدلسازی منابع آب

باوجود موفقیت‌های زیادی که SVMs در رشته‌های مختلف داشته [۷]، کاربردهای معدودی در رابطه با هیدرولوژی و مدلسازی منابع آب گزارش شده است. در ادامه مروری بر تحقیقات صورت گرفته پیرامون کاربرد ماشین‌های بردار پشتیبان در زمینه‌ی منابع آب بیان شده است.

تا آنجا که دامنه‌ی تحقیقات نشان می دهد اولین کاربردهای SVMs در زمینه منابع آب توسط دیبایک و همکارانش[۱۵] در سال ۲۰۰۱ ارائه شد [۸]. در این مطالعه به بررسی کلی تئوری یادگیری آماری (SLT) و روش ماشین‌های بردار پشتیبان پرداخته شد و سپس با استفاده از داده‌های موجود کاربرد آن در زمینه‌ی طبقه‌بندی و رگرسیون مورد بحث قرار گرفت. به عنوان یکی از کاربردهای SVMs، طبقه‌بندی نوع زمین با استفاده از عکس‌های راه دور (هوایی و ماهواره‌ای) صورت گرفته است، به طوری که با استفاده از تحلیل عکس‌ها و بر اساس اطلاعات فضایی و طیفی آنها، نوع پوشش زمین اعم از جنگل، آب، جاده و ... مشخص گردید. دیبایک برای انجام این مدلسازی از نرم‌افزار اصلاح شده دانشگاه Royal Holloway لندن و آزمایشگاه تحقیقاتی AT&T استفاده نمود. همچنین از توابع کرنل متفاوتی در این کاربرد استفاده شده که بهترین نتایج توسط تابع RBF حاصل گردید. او این نتایج را با نتایج حاصل از طبقه‌بندی توسط روش SOFM (Self Organizing Feature Map Classifier) مقایسه نمود و عملکرد بهتر SVM در مقابل SOFM را متذکر شد.

در این مقاله همچنین SVMs به منظور رگرسیون، و جهت مدلسازی بارندگی-رواناب مدنظر قرار گرفت. در بررسی امکان استفاده از آن در مدلسازی این فرایند هیدرولوژیکی، روش پیشنهادی در سه حوضه با ابعاد متفاوت و شدت بارندگی متفاوت به کار گرفته شد و عملکرد آن با روشهای دیگر از جمله ANN و مدل فیزیکی (مفهومی) به نام SMAR مقایسه گردید. در این مقاله بارندگی، تبخیر روزانه و همچنین مقدار میانگین آنها در n روز گذشته به عنوان داده‌های ورودی و مقدار دبی جریان به عنوان داده خروجی در نظر گرفته شده‌اند. دیبایک و همکارانش در این کاربرد نیز نشان دادند تابع RBF از عملکرد بهتری نسبت به سایر توابع کرنل برخوردار است. نتایج آنها نشان می‌داد که در مرحله آزمون مدل SVM به طور میانگین ۱۵ درصد دقت تخمین رواناب را در مقایسه با ANN افزایش داده است. همچنین اختلاف بسیار کم دقت نتایج در مرحله آموزش و آزمون را نشاندهنده‌ی ویژگی عمومیت‌پذیری SVMs در مقابل ANNs دانستند.

در سال ۲۰۰۲ لیونگ و همکارش[۱۶] کاربرد دیگری از روش مذکور را بررسی نمودند [۶]. در این مطالعه از داده‌های سیلابی شهر Dhaka در بنگلاش استفاده شد تا توانایی پیش‌بینی SVM در تخمین تراز سیل و در

[15] Dibike *et al.*
[16] Liong and Sivapragasam

قالب دوره‌های ۱ تا ۷ روزه، مورد بررسی قرار گیرد. به علاوه مقایسه‌ای بین نتایج این بررسی با مدلهای پیشین استفاده شده نظیر ANN صورت پذیرفت.

با توجه به خسارات زیان‌بار سیلابها، این دو محقق سعی نمودند تا در راستای کاهش خسارات، پیش‌بینی دقیق‌تری از سیل داشته باشند. داده‌های مورد استفاده آنها بارش، تراز آب رودخانه و دبی در محل‌های معین را شامل می‌شد. در ابتدا داده‌های تراز سطح آب ۸ ایستگاه مطالعه گردید و سپس با استفاده از آنالیز حساسیت، تنها ۵ ایستگاه مؤثر در پیش‌بینی تراز سیل دانسته شد. مقادیر متفاوتی از پارامتر c (یکی از پارامترهای SVM) مورد سعی و خطا قرار گرفته و همچنین مقدار γ (به عنوان پارامتر تابع کرنل RBF) نیز با استفاده از این روند تعیین گردید.

لیونگ و همکارش در این تحقیق نشان دادند که روش SVMs دارای قدرت تخمینی حداقل برابر با ANNs است. در واقع برای پیش‌بینی درازمدت، SVM نتایج بهتری نسبت به ANN در هر دو مرحله‌ی آموزش و آزمون به نمایش گذارد. به عنوان نمونه بیشینه مقدار خطای پیش‌بینی تراز آب در ۷ روز آینده برابر با مقدار ۰/۷ متر برای مدل SVM و ۰/۸۵۷ متر برای مدل ANN بوده است. همچنین در این مطالعه محقق به بررسی مزایای مدل SVM در مقایسه با ANN پرداخته و علت استفاده از این روش را توجیه نموده است.

پس از معرفی SVMs توسط دو مطالعه‌ی پیشین، تحقیقات فراوانی در زمینه مدلسازیهای هیدرولوژیکی و منابع آب آغاز گردید به طوریکه در طی سالهای ۲۰۰۴ تا ۲۰۰۶ موجی از کاربردهای متنوعی از این روش توسط محققین ارائه شد. آصفا و خلیل از جمله افرادی بودند که کاربردهای متنوعی از روش SVMs را بررسی نمودند.

در سال ۲۰۰۴ آصفا به همراه همکارانش[17] اولین مطالعه‌ی خود را در این زمینه آغاز نمود [۹]. این تحقیق روشی برای طراحی شبکه‌های بلندمدت کنترل هد آبهای زیرزمینی، بر اساس ماشین‌های بردار پشتیبان ارائه می‌داد به طوری که چاه‌های مشاهده‌ای اضافی (که در تخمین سطح پیزومتریک آبخوان مؤثر نیستند) کاهش یابد.

به طور کلی شبکه‌های کنترل آبهای زیرزمینی به دو دسته تقسیم می‌گردند که شامل (۱) شبکه‌های کنترل آلودگی آبهای زیرزمینی و (۲) شبکه‌های مشاهده‌ی هد آبهای زیرزمینی می‌باشند. در دسته‌ی دوم شبکه‌های کنترل بلندمدت آندسته از چاه‌هایی هستند که از زیر مجموعه‌ای از دیگر چاه‌های موجود انتخاب می‌شوند تا کنترل مرتب و دوره‌ای (ماهانه یا فصلی) از تغییرات هد در آنها مشاهده و بررسی گردد. با توجه به این مسئله، در این مطالعه برای انتخاب این چنین چاه‌ها، روش SVMs ارائه شده تا بتواند آندسته از چاه‌های مشاهده‌ای که در توصیف سطح پیزومتریک آبخوان نقش مهمتری دارا هستند را برگزینند. علاوه بر این،

[17] Asefa *et al.*

تشخیص یک زیر مجموعه از کلیه‌ی چاه‌های مشاهده‌ای و کنترل سطح پیزومتری در آنها به طور منظم باعث عدم نیاز به کنترل هد در تمامی چاه‌ها و نهایتاً کاهش هزینه‌ها مربوطه خواهد شد بدون آنکه در دقت کنترل خللی حاصل گردد. آصفا در این مطالعه روشهای دیگری را نیز برای این منظور بررسی نمود و مشخص شد که این روشها از لحاظ محاسباتی پرهزینه و وقت‌گیر هستند. او هد آب در آبخوان را به صورت تابع هدف مدل انتخاب نمود و بردارهای پشتیبان را که در روند مدلسازی SVM ایجاد می شوند، به عنوان چاه‌های کنترل بلندمدت در نظر گرفت. داده‌های ورودی شامل موقعیت چاه، و میزان هد در آن به عنوان خروجی در نظر گرفته شد.

در بررسی رابطه‌ی بین اندازه شبکه کنترل (تعداد چاه های انتخابی) و مقدار پارامتر ε مدل SVM نشان داده شد که افزایش میزان دقت (و یا کاهش مقدار پارامتر ε) منجر به افزایش تعداد چاه‌های موردنیاز برای کنترل (بردارهای پشتیبان) خواهد شد. به عنوان نمونه برای ε=۰/۱ تعداد ۲۳ چاه و برای ε=۰/۰۵ تعداد ۶۵ چاه از بین ۳۵۰ چاه نیاز بود. نتیجه دیگر اینکه، موقعیت چاه‌های انتخاب شده بیانگر مناطقی بودند که سطح پیزومتریک بیشترین عدم قطعیت را داشته و نزدیک بودن دو چاه به یکدیگر تغییرات قابل توجه هد در آن منطقه را بیان می نمود. در این مطالعه مقادیر بهینه وراپارامترها توسط عمل تصدیق متقاطع تعیین شده‌اند و تابع کرنل مورد استفاده RBF بوده است.

در مطالعه‌ی مشابه دیگری که توسط این محقق در سال ۲۰۰۵ صورت گرفت [۱۰]، سعی شد تا با استفاده از SVMs به طراحی شبکه‌های کنترل آلودگی آبهای زیرزمینی پرداخته شود. این مسئله همانند کاربرد پیشین SVMs (طراحی شبکه‌های کنترل هد) بود با این تفاوت که این مطالعه در زمینه‌ی شبکه‌های کنترل نشت آلودگی صورت گرفت. این شبکه‌ها وظیفه‌ی نظارت بر نشت آلودگی از منشایی خاص (مانند Landfill) و در یک مرز مشخص را دارا هستند. هدف این مقاله یافتن چنین شبکه‌ای است به طوری که بتواند بیشترین احتمال یافتن نشت (شبکه‌ای با بالاترین میزان اطمینان پذیری) را با کمترین هزینه برآورده سازد. برای این منظور تعداد ده عدد چاه انتخاب شده تا بتوان از میان آنها شبکه‌ی بهینه را انتخاب نمود.

در استفاده از SVM، محل چاه‌های کنترل به عنوان ورودی، و میزان اطمینان پذیری شبکه به عنوان خروجی در نظر گرفته شده است. میزان اطمینان پذیری هر شبکه توسط مدلسازی فیزیکی که بر روی یک منطقه‌ی نمونه و توسط نرم‌افزارهای ModFlow و ModPath انجام شد، محاسبه گردید.

توابع کرنل مورد استفاده در این تحقیق SRF[۱۸] و Spline بوده‌اند. محقق عملکرد مدل SVM را با مقدار c در بازه ۱۰ تا ۱۰۰ (برای تابع کرنل SRF) بهینه دانسته و لذا مقدار c برابر با ۹۰ و مقدارε=۰/۰۰۱ در نظر گرفته شده است. آصفا نتایج بدست آمده از این روش را با نتایج حاصل از مدلسازی توسط ANN مقایسه نمود و بهبود عملکرد SVM را نسبت به ANN مشخص ساخت. او همچنین با مقایسه‌ی نتایج با مدل فیزیکی

[18] Strongly Regularized Fourier