مجله سلامت و محیط، فصلنامه‌ی علمی پژوهشی
انجمن علمی بهداشت محیط ایران
دوره ششم، شماره اول، بهار ۱۳۹۲، صفحات ۱ تا ۱۰

ارایه مدلی مناسب با استفاده از ماشین بردار پشتیبان برای پیش‌بینی غلظت روزانه مونوکسیدکربن در هوای شهر تهران

روح‌اله نوری[۱]، غلامعلی هشیاری‌پور[۲]، خسرو اشرفی[۳]، عمران راستی[۴]

دریافت: ۹۰/۱۱/۱۹ پذیرش: ۹۱/۰۲/۱۷

چکیده

زمینه و هدف: *پیش‌بینی دقیق آلاینده‌های هوا، به عنوان اولین گام جهت برخورد مناسب با مشکل آلودگی هوا، می‌تواند اطلاعات مفیدی را برای برنامه‌ریزی جهت مقابله با این موضوع در اختیار مدیران ذیربط قرار دهد. در این مقاله با توجه به معضل آلاینده مونوکسیدکربن (CO) در هوای شهر تهران، اقدام به ارایه مدلی مناسب برای پیش‌بینی این آلاینده شده است.*

روش بررسی: *برای این منظور از اطلاعات آلاینده‌های هوا و پارامترهای هواشناسی ثبت شده در ایستگاه قلهک در شمال تهران که مجموعا ۱۲ ورودی به مدل برای پیش‌بینی غلظت میانگین روزانه CO را تشکیل می‌دادند، استفاده گردید. در گام اول این مطالعه از مدل ماشین بردار پشتیبان (SVM) برای مدل‌سازی غلظت روزانه CO استفاده شد. در گام بعد با استفاده از تکنیک انتخاب پیشرو، تعداد ورودی به مدل SVM از ۱۲ به ۷ متغیر کاهش و سپس مدل مناسبی (مدل FS-SVM) جهت پیش‌بینی غلظت روزانه CO توسعه داده شد.*

یافته‌ها: *به منظور ارزیابی دقت مدل‌های SVM و FS-SVM در پیش‌بینی روزانه CO در شهر تهران از شاخص ضریب همبستگی استفاده گردید. ضریب همبستگی در مرحله تست هر دو مدل مذکور تقریبا یکسان بوده وحاکی از دقت مناسب هر دو مدل در پیش‌بینی روزانه CO است. به هر حال باید توجه داشت که استفاده از مدل FS-SVM به دلیل کاهش تعداد متغیرهای ورودی نسبت به مدل SVM، با هزینه‌های کمتر محاسباتی و اقتصادی همراه است.*

نتیجه‌گیری: *نتایج به دست آمده از این دو مدل مشخص نمود که اگرچه هر دو مدل از دقت تقریبا یکسانی در پیش‌بینی غلظت روزانه CO برخوردارند، اما مدل FS-SVM به دلیل نیاز به تعداد کمتر ورودی و در نتیجه حجم محاسباتی کمتر، می‌تواند از عملکرد بهتری در این زمینه برخوردار باشد.*

واژگان کلیدی: مونوکسیدکربن، ماشین بردار پشتیبان، انتخاب پیشرو، تهران

۱- **(نویسنده مسئول)**: دکترای محیط زیست، استادیار دانشکده محیط زیست، دانشگاه آزاد اسلامی واحد ملارد *roohollahnoori@gmail.com*
۲ - دانشجوی دکترای ژئوفیزیک، موسسه ژئوفیزیک، دانشگاه هامبورگ، آلمان
۳- دکترای محیط زیست، استادیار دانشکده تحصیلات تکمیلی محیط زیست، دانشگاه تهران
۴- دکترای جغرافیای سیاسی، استادیار دانشکده جغرافیا، دانشگاه بیرجند

مقدمه

امروزه مشکلات زیست‌محیطی و سلامتی ناشی از آلودگی هوا در کلان شهرها به یک چالش اساسی تبدیل شده است. این مشکل در مورد شهر تهران به دلیل حجم ترافیکی سنگین ناشی از ترابری، استفاده از خودروهای غیراستاندارد، احتراق ناقص سوخت‌های مورد استفاده خودروها و بی‌توجهی که طی سال‌های گذشته نسبت به آلودگی هوا در این شهر صورت گرفته، از اهمیت ویژه‌ای برخوردار است (۱). پیش‌بینی غلظت روزانه آلاینده‌های هوا اولین گام اساسی در برنامه‌ریزی کاهش اثرات آنها است. برای این منظور تاکنون روش‌های زیادی برای پیش‌بینی غلظت آلاینده‌های هوا ارایه شده است که آنها را می‌توان به دو دسته روش‌های قطعی و آماری تقسیم نمود. مدل‌های قطعی (Deterministic Models) آلودگی هوا که اساسا حالت پایه انتقال آشفتگی در اتمسفر را منعکس می‌کنند، به عنوان ابزاری خبره جهت مدل‌سازی آلاینده‌های گازی و ذرات به شمار می‌روند؛ اما نتایج آنها همیشه توسط مقدار قابل توجهی خطا تحت تاثیر قرار می‌گیرد. این امر می‌تواند به دلیل تشریح جزیی و مختصر پروسه‌های پیچیده اتمسفر در این مدل‌ها باشد. فاکتورهای زیادی در افزایش خطای این مدل‌ها تاثیر داشته که از مهم‌ترین آنها عدم قطعیت ناشی از تغییرپذیری ذاتی اتمسفر است. از طرفی تمرکز چنین مدل‌هایی بر این فرض استوار است که آلاینده‌ها در شرایط همگنی پخش می‌شوند، اما عملا وجود زمین می‌تواند عاملی مهم در ناهمگنی آشفتگی در مسیر عمودی باشد. علاوه بر این ورودی مدل‌های مذکور، که غالبا از نوع گوسی هستند، اغلب بر مبنای طرح‌ریزی ساده‌ای بنا شده‌اند که آشفتگی را در کلاس‌های پایداری فرض می‌کنند و این در حالیست که هر کلاس بازه وسیعی از شرایط پایداری اتمسفر را پوشش می‌دهد و به مکانی که در آن ارزیابی می‌شود بستگی دارد (۲). این عوامل در کنار عواملی دیگر نظیر مشکلات دسترسی به ضرایب انتشار آلاینده‌ها، که دسترسی به آنها در بیشتر موارد با مشکلاتی همراه است و همچنین مشکل بودن ساختار مدل‌های قطعی باعث شده تا طی سال‌های اخیر توجه خاصی به مدل‌های پیشرفته آماری شود (۱). روش‌های آماری با استفاده از داده‌های موجود هواشناسی و آلودگی و تحلیل ارتباط آماری بین آنها، راه‌کارهای ساده‌تری برای پیش‌بینی غلظت آلاینده‌ها به شمار می‌روند و تحقیقات صورت گرفته نیز در زمینه پیش‌بینی کوتاه مدت آلاینده‌های هوا با استفاده از این روش‌ها، سودمندی آنها را به اثبات رسانده است (۳و۴). همچنین روش‌های آماری علاوه بر این که به اطلاعات انتشار و ضرایب انتشار نیازی ندارند، از ساختاری ساده‌تر نیز نسبت به مدل‌های قطعی برخوردارند. تا به حال روش‌های آماری متعددی برای پیش‌بینی غلظت آلاینده‌های هوا مورد استفاده قرار گرفته‌اند که در این راستا می‌توان به مدل‌های رگرسیون خطی و غیرخطی (۱و۵)، شبکه عصبی مصنوعی (۶و۷) و سیستم استنتاج تطبیقی عصبی ـ فازی (۷) اشاره نمود. همچنین استفاده از دیگر روش‌های آماری مانند ماشین بردار پشتیبان (Support Vector Machine-SVM) که اولین بار توسط وپنیک (۱۹۹۵) ریاضیدان روسی ارایه شد، در پژوهش‌های مربوط به مسایل زیست‌محیطی و به تبع آن امر آلودگی هوا طی چند سال اخیر مورد توجه برخی از محقیق قرار گرفته است. Noori و همکاران (۸) طی تحقیقی با استفاده از مدل SVM اقدام به پیش‌بینی میزان هفتگی زباله شهر تهران نمودند. همچنین Noori و همکاران (۹) برای پیش‌بینی ضریب انتشار طولی در رودخانه‌های طبیعی از SVM استفاده نموده و نتایج این مدل را در مقایسه با مدل‌های کلاسیک رگرسیونی بهتر گزارش کردند. Lu(۱۰) در تحقیقی به مقایسه SVM و شبکه عصبی تابع پایه شعاعی (Radial Base Function-RBF) برای مدل‌سازی کیفی هوای شهر هنگ کنگ در چین پرداختند. آنها در نهایت برتری مدل SVM را نسبت به مدل RBF گزارش نمودند. Osowski (۱۱) مدل ترکیبی SVM با تبدیل موجک (Wavelet) را برای پیش‌بینی پارامترهای کیفی هوا پیشنهاد کردند. Salazar و همکاران (۱۲) برای پیش‌بینی ماکزیمم غلظت روزانه ازن تروپوسفریک در ناحیه‌ای از کشور امریکا از مدل SVM استفاده نمودند. در حالت کلی نتایج به دست آمده از مدل SVM در زمینه پیش‌بینی آلودگی هوا امیدوارکننده بوده و روزبه‌روز بهبودهایی در این زمینه توسط محققین مختلف ارایه می‌شود که می‌تواند در کشور ایران نیز با برنامه‌ریزی صحیح از این ابزار قدرتمند، راه‌کارهای مناسبی

جهت مدیریت آلودگی هوا در اختیار مدیران ذیربط قرار گیرد. در همین راستا و در تحقیق مذکور جهت بررسی عملکرد مدل SVM در پیش‌بینی آلودگی هوای شهر تهران و با توجه به اهمیت آلاینده گازی CO در این شهر، پیش‌بینی مقدار میانگین روزانه این آلاینده با استفاده از اطلاعات هواشناسی و آلودگی هوا در ایستگاه قلهک مدنظر قرار گرفته است. همچنین تاثیر عملکرد الگوریتم انتخاب پیشرو (Forward Selection-FS) در گزینش ورودی به مدل SVM نیز از اهداف دیگر این تحقیق است.

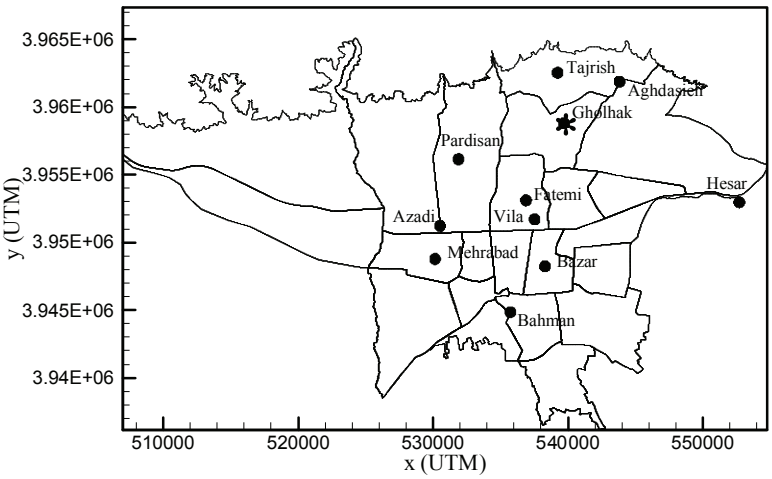
منطقه مورد مطالعه و اطلاعات مساله

شهر تهران در کوه‌پایه‌های جنوبی رشته کوه البرز گسترده شده و جمعیت آن مطابق با آخرین آمارگیری حدود ۸ میلیون نفر برآورد گردیده است. نتایج به دست آمده از تحقیقی در مورد آلاینده‌های هوا در این شهر، بیانگر این واقعیت است که ۸۹٪ وزن کل آلاینده‌های هوای شهر تهران از خودروها منتشر شده و ۱۱٪ مابقی مربوط به منابع ثابت است. آلاینده CO نسبت به بقیه آلاینده‌های هوا در شهر تهران از اهمیت بیشتری برخوردار است، به طوری‌که بیش از سه‌چهارم وزن آلاینده‌های هوا در این شهر را آلاینده مذکور تشکیل می‌دهد (۱۳). در شهر تهران مطابق شکل ۱ یازده ایستگاه سنجش آلودگی هوا وجود دارد که کار اندازه‌گیری غلظت آلاینده‌های شاخص هوا را انجام می‌دهند. در این تحقیق به منظور پیش‌بینی میانگین غلظت روزانه CO در یک روز بعد، از داده‌های هواشناسی شامل دمای هوا (T)، رطوبت نسبی (Hum)، فشار هوا (Press)، سرعت باد (WS)، جهت باد (WD)، تابش خورشیدی (Solar) و داده‌های آلودگی هوا شامل دی‌اکسیدگوگرد (SO_2)، کل هیدروکربن‌ها (THC)، ازن (O_3)، اکسیدهای نیتروژن (NO_X) و ذرات معلق با قطر معادل یا کمتر از ۱۰ μ (PM_{10}) در ایستگاه قلهک واقع در شمال تهران، ثبت شده در سال‌های ۱۳۸۳ و ۱۳۸۴ استفاده گردیده است. قابل ذکر است که به علت خاموشی و مشکلات فنی دستگاه سنجش آلودگی هوای ایستگاه قلهک، اطلاعات برخی از روزها در طی این دو سال در دسترس نبوده و مجموعا بعد از مرتب کردن اطلاعات ثبت‌شده، از اطلاعات ۴۸۳ روز ثبت شده در ایستگاه مذکور در طی این دو سال استفاده شده است.

مواد و روش‌ها

انتخاب پیشرو

یکی از مهم‌ترین مسایل در توسعه مدل‌های آماری، پیش‌پردازش اطلاعات ورودی به منظور انتخاب ورودی‌های مناسب به مدل است؛ زیرا بعضا برخی از اطلاعات نه تنها اثر مستقیم چندانی بر خروجی مدل ندارند، بلکه باعث کاهش اثر سایر داده‌ها نیز خواهند شد که می‌تواند به یک مدل‌سازی نامناسب منجر گردد. همچنین بخشی از داده‌ها ممکن است با



شکل ۱: موقعیت ایستگاه قلهک (*) در بین ایستگاه ددیگر در شهر تهران

یکدیگر همبستگی نشان دهند و استفاده از تمام آنها در مدل صرفا به افزایش ابعاد مدل و حجم محاسبات آن منجر شود. بنابراین بهتر است در برخورد با این گونه اطلاعات، موثرترین آنها انتخاب شده و سایر داده‌های دارای همبستگی با آن، از ورودی حذف شود که علاوه بر بهینه‌سازی جواب‌های مدل، حجم محاسبات آن را نیز کاهش‌دهد. روش‌های متعددی به منظور پیش‌پردازش داده‌ها و یا انتخاب ورودی مدل پیشنهاد شده (۱۶-۱۴) که در این تحقیق از الگوریتم FS استفاده شده است. این روش به طور موفقیت‌آمیزی توسط محققین زیادی مورد استفاده قرار گرفته است (۱۷و۱۸). جزییات بیشتر در مورد FS درمنابع دیگر در دسترس است (۱۸).

ماشین بردار پشتیبان

با توجه به تعداد زیاد کتب و مقالات موجود در زمینه مباحث تئوری مدل SVM (۲۱-۱۹) در این مقاله تنها به شرح خلاصه‌ای از مدل مورد استفاده SVM به نام SVM -ε اکتفا می‌شود. 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 رابطه‌ای مانند (y=f(x)+noise)تعریف شود. بنابراین موضوع اصلی پیدا کردن فرم تابع f است که بتواند به صورت صحیح الگوهای جدیدی را که SVMتاکنون تجربه نکرده است، پیش‌بینی کند. این تابع به وسیله آموزش مدل SVM بر روی یک مجموعه داده به عنوان مجموعه آموزش که شامل پروسه‌ای جهت بهینه‌سازی دائمی تابع خطاست، قابل دسترسی است. در این مطالعه مدل SVM-ε دلیل کاربرد گسترده آن در مسایل رگرسیونی استفاده گردیده است. برای این مدل تابع خطا به صورت زیر تعریف می‌شود:

$$\frac{1}{2}w^T w + C\sum_{i=1}^{N}\xi_i + C\sum_{i=1}^{N}\xi_i^{\bullet} \qquad (۱)$$

تابع خطای مذکور لازم است که با توجه به محدودیت‌های زیر کمینه گردد:

$$\begin{aligned} & w^T\phi(x_i)+b-y_i \le \varepsilon+\xi_i^{\bullet} \\ & y_i - w^T\phi(x_i)-b \le \varepsilon+\xi_i^{\bullet} \\ & \xi_i,\xi_i^{\bullet} \ge 0 \quad , \quad i=1,...,N \end{aligned} \qquad (۲)$$

در روابط بالا، C ثابت گنجایش (Capacity)، w بردار ضرایب، w^T ترانهاده بردار ضرایب،ξ_iو $\xi_i^{\bullet}$ ضرایب کمبود (Slack)، b ضریبی ثابت وN الگوهای آموزش مدل و نهایتا Ø تابع کرنل است. تابع کرنل RBF به عنوان بهترین انتخاب از بین دیگر توابع کرنل گزارش شده است (۲۲). بنابراین در این تحقیق، تابع مذکور که توسط γ در معادله (۳) تعریف گردیده به کار گرفته شده است.

$$K(x_i,x)=\exp\left(-\gamma\left|x_i-x\right|^2\right) \qquad (۳)$$

با توجه به معادلات ۳-۱ لازم است برای پیشبینی میانگین غلظت روزانه CO توسط مدل SVM، سه پارامتر ε ، γ و C بهینه شوند.

بحث

توسعه مدل SVM

در این تحقیق برای پیش‌بینی میانگین غلظت روزانه CO توسط مدل SVM، از نرم افزار STATISTICA استفاده گردید. همان طور که قبلا نیز ذکر شد، جهت مدل‌سازی میانگین غلظت روزانه CO توسط مدل SVM با توجه به نوع تابع کرنل مورد استفاده در این تحقیق (تابع کرنل RBF) لازمست سه پارامتر ε، γ و C بهینه شوند. مقدار پیشنهادی مدل SVM برای پارامتر γ معادل با ۱/K است که در این رابطه K معادل تعداد متغیرهای ورودی به مدل است (در این مرحله از تحقیق K=۱۲ است). دو پارامتر ε و C نیز با استفاده از یک برنامه بهینه‌سازی مناسب لازمست است بهینه شوند. روش پیش فرض در نرم افزار STATISTICA برای بهینه‌سازی ε و C روش جستجوی شبکه (Grid Search) است (۲۳). با توجه به مطالب ذکر شده، در این مرحله از مطالعه مقدار γ معادل با ۰/۰۸۳ قرار داده شد. دو پارامتر ε و C نیز توسط روش جستجوی شبکه بهینه شدند. برای بهینه‌سازی این دو پارامتر، محدوده تغییرات ε از ۰/۰۰۱ تا ۰/۵ با گام ۰/۰۰۱ و محدوده تغییرات C نیز از ۱ تا ۲۰۰ با گام ۵ انتخاب شدند. بدین‌طریق مقادیر بهینه ε و C به ترتیب معادل ۰/۱۷۸ و ۱۴۶ محاسبه گردیدند. مقدار ضریب همبستگی (R) نیز در مراحل آموزش و تست مدل SVM برای پیش‌بینی میانگین غلظت روزانه CO به ترتیب برابر با ۰/۸۲ و ۰/۸۸ به دست آمد.

جدول ۱: همبستگی تکتک متغیرهای ورودی با خروجی مدل (غلظت CO)

متغیر	R^2	متغیر	R^2
T	۰/۲۶۹۹	O_3	۰/۰۲۶۷
Hum	۰/۲۰۰۲	WS	۰/۰۲۶۶
Press	۰/۱۸۲۰	NO_x	۰/۰۲۲۹
Solar	۰/۰۹۱۰	THC	۰/۰۱۹۸
SO_2	۰/۰۷۸۲	CH_4	۰/۰۱۴۱
PM_{10}	۰/۰۷۰۷	WD	۰/۰۰۷۷

توسعه مدل ترکیبی SVM با FS (مدل FS-SVM)

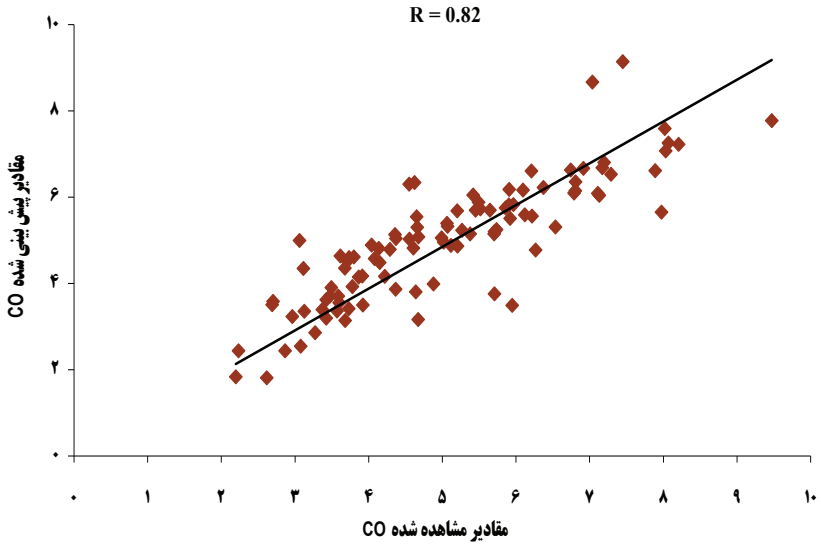
در این مقاله جهت انتخاب ورودی‌های مدل، از روش FS استفاده شده‌است. شایان ذکر است کلیه ورودی‌ها برای اعمال FS بایستی استانداردسازی شود به نحوی که ابعاد تمامی متغیرها به بازه [۱و ۱-] منتقل گردد. نتایج مربوط به FS در جدول‌های ۱ و ۲ آمده است. از جدول ۱ ملاحظه می‌شود که دما (T) دارای بیشترین همبستگی است و به عنوان اولین ورودی مدل انتخاب شده است. جدول ۲ نشان‌دهنده نتایج فرایند FS است.

از این جدول می‌توان دریافت که با ورود متغیرهای جدید، R^2 افزایش می‌یابد، تا جایی که تغییرات آن در دو گام متوالی از ۵٪ کمتر شده و قابل صرف نظر کردن است. بنابراین متغیرهای نظیر R^2= ۰/۶۳۲۵ به عنوان ورودی‌های مدل انتخاب می‌شوند که به ترتیب عبارتند از T، SO_2، CH_4، Press، THC، NO_X و در نهایت O_3. پس از انتخاب ۷ متغیر مذکور از بین ۱۲ متغیر اصلی به عنوان بهترین ترکیب ورودی به مدل SVM، همانند مرحله قبل مدل SVM مناسب با ترکیب مذکور جهت پیش‌بینی

جدول ۲: نتایج فرایند انتخاب پیشرو

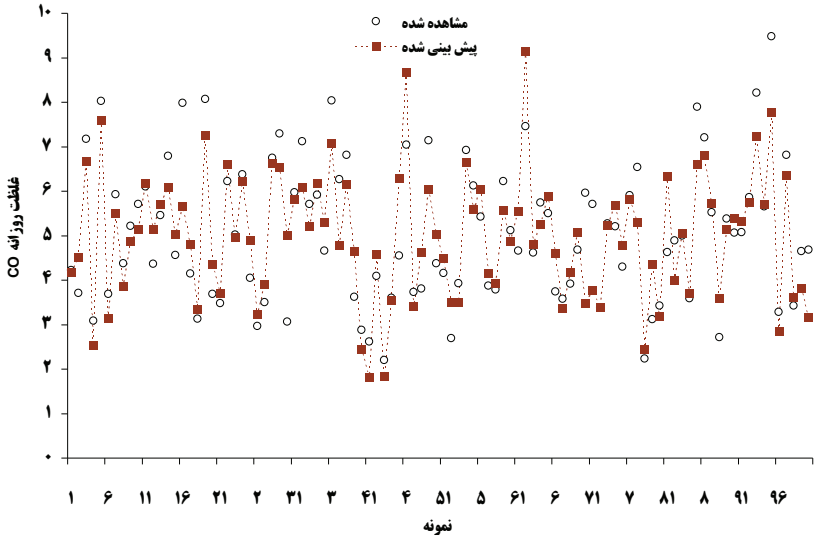
مجموعه متغیرهای ورودی	R^2
T	۰/۲۶۹۹
T, SO_2	۰/۳۴۵۸
T, SO_2, THC	۰/۴۵۹۴
T, SO_2, THC, Press	۰/۴۶۱۵
T, SO_2, THC, Press, CH_4	۰/۵۱۹۳
T, SO_2, THC, Press, CH_4, NO_x	۰/۶۰۵۵
T, SO_2, THC, Press, CH_4, NO_x, O_3	۰/۶۳۲۵×
T, SO_2, THC, Press, CH_4, NO_x, O_3, Solar	۰/۶۳۳۷
T, SO_2, THC, Press, CH_4, NO_x, O_3, Solar, PM_{10}	۰/۶۳۳۹
T, SO_2, THC, Press, CH_4, NO_x, O_3, Solar, PM_{10}, Hum	۰/۶۳۴۵
T, SO_2, THC, Press, CH_4, NO_x, O_3, Solar, PM_{10}, Hum, WD	۰/۶۳۴۷
T, SO_2, THC, Press, CH_4, NO_x, O_3, Solar, PM_{10}, Hum, WD, WS	۰/۶۳۴۹

× بعد از این مقدار، ملاحظه می‌شود که تغییر چندانی در R^2 رخ نمی‌دهد. لذا متغیرهای متناظر با این مقدار به عنوان ورودی‌های مدل انتخاب می‌شوند.



شکل ۲: ضریب همبستگی مرحله تست مدل FS-SVM

غلظت میانگین روزانه CO توسعه داده شد. مقدار γ با توجه به تعداد ورودی‌ها معادل ۰/۱۴۳ انتخاب گردید. پارامترهای ε و C نیز با استفاده از روش جستجوی شبکه به ترتیب ۰/۰۴۰ و ۱۴۶ محاسبه شدند. مقدار R در مرحله تست مدل FS-SVM معادل ۰/۸۲ به دست آمد. مقدار R مرحله آموزش مدل FS-SVM نیز معادل ۰/۸۷ محاسبه شد. با مقایسه نتایج به دست آمده از دو مدل SVM و FS-SVM مشخص می‌شود که عملکرد این دو مدل در پیش‌بینی میانگین غلظت روزانه CO در شهر تهران یکسان است و هر دو مدل از دقت تقریبا یکسانی در این مورد برخوردارند. اما با توجه به تعداد ورودی‌ها به هر یک از این دو مدل، مشخص است که مدل FS-SVM از عملکرد بهتری برخوردار است؛ زیرا با تعداد اطلاعات مورد نیاز کمتر همان دقت مدل SVM را داشته و از طرف دیگر در مدل‌های آماری تعداد متغیرهای زیاد می‌تواند باعث ناپایداری مدل مورد استفاده شود. به علاوه، تعداد ورودی کمتر باعث کاهش ابعاد مدل و درنتیجه کاهش



شکل ۳: سری زمانی مربوط به مرحله تست مدل FS-SVM

بار محاسباتی آن می‌شود. در نهایت در این تحقیق با توجه به مطالب مذکور مدل FS-SVM به عنوان مدل بهینه برای پیش‌بینی میانگین غلظت روزانه CO در هوای شهر تهران انتخاب می‌شود. شکل‌های ۲ و ۳ نتایج مرحله تست مدل FS-SVM را نشان می‌دهند. ذکر این نکته نیز لازمست که با مقایسه نتایج به دست آمده از این تحقیق با تحقیقات مشابه انجام گرفته خصوصا در شهر تهران (۱و۷)، مشخص می‌شود که عملکرد مدل SVM از مدل‌های آماری کلاسیک مانند رگرسیون خطی چند متغیره برتر بوده و با اندکی اختلاف قابل ارزیابی با مدل‌های هوشمند شبکه عصبی مصنوعی و سیستم استنتاج فازی تطبیقی است.

نتیجه گیری

هدف این مقاله پیشنهاد راهکاری جدید برای کاهش آلودگی هوا نبوده چرا که بسیاری از قانون‌ها، مصوبات و راهکارهای پیشنهادی طرح‌ها و پژوهش‌ها به علت نبود مدیریت یکپارچه شهری و نبود متولی واحد در خصوص آلودگی هوا در کاربست عملیاتی با مشکل و نقص روبرو شده‌اند. همچنین پرواضح است که حل معضل آلودگی هوای شهر تهران را در کوتاه مدت نمی‌توان انتظار داشت. یافته‌های این تحقیق نیز بیشتر به مدیریت بحران آلودگی هوا و پیشبینی شرایط اضطرار و بحران کمک می‌کند و قادر است ۲۴ ساعت قبل با احتمال نزدیک به یقین وضعیت آلودگی هوا را در شبانه روز آینده پیش‌بینی نماید. به همین منظور هدف اصلی در مقاله مذکور ارایه چارچوبی مناسب با استفاده از دو تکنیک ماشین بردار پشتیبان و انتخاب پیشرو جهت پیش‌بینی غلظت یکی از مهم‌ترین آلاینده‌های هوا در شهر تهران یعنی CO قرار داده شد. در همین راستا دو مدل SVM و FS-SVM با استفاده از اطلاعات آلاینده‌های هوا و پارامترهای هواشناسی توسعه داده شدند. جهت توسعه این دو مدل ابتدا پارامترهای آنها توسط برنامه بهینه‌سازی جستجوی شبکه و توصیه‌های انجام شده در این زمینه بهینه و سپس نتایج مراحل آموزش و تست دو مدل ارایه گردیدند. نتایج این تحقیق در حالت کلی حاکی از برتری عملکرد مدل FS-SVM بود که با تعداد ورودی کمتر توانسته است همان دقت مدل SVM را حفظ نماید. در پایان متدولوژی ارایه شده در این تحقیق می‌تواند برای پیش‌بینی دیگر آلاینده‌های مهم شهر تهران مانند ذرات معلق با قطر معادل یا کمتر از ۱۰ μ m نیز مورد استفاده قرار گیرد. علاوه بر این به متولیان معضل آلودگی هوا و مدیریت شهری کمک خواهد نمود تا تصمیم‌گیری مناسب و به موقعی را در خصوص هشدار و اطلاع‌رسانی به موقع به مردم و یا اقدامات دیگر از جمله تعطیلی مدارس و ادارات و اعمال محدودیت در تردد خودروها صورت دهند. همچنین نتیجه این کار علمی و پژوهشی می‌تواند به اختلاف نظر متولیان این امر درخصوص اعلام شرایط اضطرار نیز پایان دهد.

منابع

1. Noori R, Ashrafi K, Azhdarpour A. Comparison of ANN and PCA based multivariate linear regression applied to predict the daily average concentration of CO: a case study of Tehran. Journal of the Earth Space Physics. 2008;34(1):135-52 (in Persian).
2. Pelliccioni A, Tirabassi T. Air dispersion model and neural network: a new perspective for integrated models in the simulation of complex situations. Environmental Modelling & Software. 2006;21(4):539-46.
3. Gilbert RO. Statistical Methods for Environmental Pollution Monitoring. New York: John Wiley & Sons, Inc; 1987.
4. Zannetti P. Air Pollution Modelling: Theories, Computational Methods and Available Software. New York: van Nostrand; 1990.
5. Nunnari G, Dorling S, Schlink U, Cawley G, Foxall R, Chatterton T. Modelling SO2 concentration at a point with statistical approaches, Environmental Modelling & Software. 2004;19(10):887–905.
6. Gardner MW, Dorling SR. Neural network modelling and prediction of hourly NOx and NO2 concentrations in urban air in London. Atmospheric Environment. 1999;33(5):709–19.
7. Noori R, Hoshyaripour G, Ashrafi K, Nadjar-Araabi B. Uncertainty analysis of developed ANN and ANFIS models in prediction of carbon monoxide daily concentration. Atmosphric Environment. 2010;44(4):476-82.
8. Noori R, Abdoli MA, Ameri- Ghasrodashti A, Jalili-Ghazizade M. Prediction of municipal solid waste generation with combination of support vector machine and principal component analysis: a case study of Mashhad. Environmental Progress & Sustainable Energy. 2009;28(2):249-58.
9. Noori R, Karbassi A, Farokhnia A, Dehghani M. Predicting the longitudinal dispersion coefficient using support vector machine and adaptive neuro-fuzzy inference system techniques. Environmental Engineering Science. 2009;26(10):1503-10.
10. Lu WZ, Wang WJ. Potential assessment of the support vector machine method in forecasting ambient air pollutant trends. Chemosphere. 2005;59:693-701.
11. Osowski S, Garanty K. Forecasting of the daily meteorological pollution using wavelets and support vector machine. Engineering Applications of Artificial Intelligence. 2007;20:745-55.
12. Salazar-Ruiz E, Ordieres JB, Vergara EP, Capuz-Rizo SF. Development and comparative analysis of tropospheric ozone prediction models using linear and artificial intelligence-based models in Mexicali, Baja California (Mexico) and Calexico, California (US). Environmental Modelling & Software. 2008;23:1056-69.
13. Bayat R. Source Apportionment of Tehran's air pollution [dissertation]. Tehran: Sharif University of Technology; 2005 (in Persian).
14. Wang XX, Chen S, Lowe D, Harris CJ. Sparse support vector regression based on orthogonal forward selection for the generalised kernel model. Neurocomputing. 2006;70:462-74.
15. Noori R, Karbassi AR, Sabahi MS. Evaluation of PCA and Gamma test techniques on ANN operation for weekly solid waste prediction. Journal of Environmental Management. 2010;91:767-71.
16. Noori R, Abdoli MA, Jalili-Ghazizade M, Samieifard R. Comparison of ANN and PCA based multivariate linear regression applied to predict the weekly municipal solid waste generation in Tehran. Iranian Journal of Public Health. 2009;38:74-84.
17. Chen S, Billings SA, Luo W. Orthogonal least squares methods and their application to non-linear system identification. International Journal of Control. 1989;50(5):1873-96.
18. Khan JA, Aelst SV, Zamar RH. Building a robust linear model with forward selection and stepwise procedures. Computational Statistics & Data Analysis. 2007;52(1):239-48.
19. Vapnik VN. The Nature of Statistical Learning Theory. 2nd ed. New York: Springer-Verlag; 1995.
20. Yu PS, Chen ST, Chang IF. Support vector regression for real-time flood stage forecasting. Journal of Hydrology. 2006;328(3-4):704-16.
21. Chen ST, Yu PS. Pruning of support vector networks on flood forecasting. Journal of Hydrology. 2007;347(1-2):67-78.
22. Dibike YB, Velickov S, Solomatine D, Abbott MB. Model induction with support vector machines: introduction and applications. Journal of Computing in

Civil Engineering. 2001;15(3):208-16.
23. Hsu CW, Chang CC, Lin CJ. A Practical guide to support vector machine classification. University of California: Irvine Press; 2003 [cited 2011 Jun 25]. Available from: http:// www.csie.ntu.edu.tw/~cjlin/ papers/guide/guide.pdf.

Introducing an Appropriate Model using Support Vector Machine for Predicting Carbon Monoxide Daily Concentration in Tehran Atmosphere

***Roohollah Noori[1], Gholamali Hoshyaripour[2], Khosro Ashrafi[3], Omran Rasti[4]**

[1]Department of Civil Engineering, Islamic Azad University, Malard Branch, Tehran, Iran

[2]Department of Geophysics, Institute of Geophysics, University of Hamburg, Hamburg, Germany

[3]Department of Environmental Engineering, Graduate Faculty of Environment, University of Tehran, Tehran, Iran

[4]Department of Politic Geography, Faculty of Geography, University of Birjand, Khorasan Jonoubi, Iran

Received; 19 February 2012 Accepted; 17 May 2012

ABSTRACT

Backgrounds and Objectives: Precise air pollutants prediction, as the first step in facing air pollution problem, could provide helpful information for authorities in order to have appropriate actions toward this challenge. Regarding the importance of carbon monoxide (CO) in Tehran atmosphere, this study aims to introduce a suitable model for predicting this pollutant.

Materials and Method: We used the air pollutants and meteorological data of Gholhak station located in the north of Tehran; these data provided 12 variables as inputs for predicting the average CO concentration of the next day. First, support vector machine (SVM) model was used for forecasting CO daily average concentration. Then, we reduced the SVM inputs to seven variables using forward selection (FS) method. Finally, the hybrid model, FS-SVM, was developed for CO daily average concentration forecasting.

Result: In the research, we used correlation coefficient to evaluate the accuracy of both SVM and FS-SVM models. Findings indicated that correlation coefficient for both models in testing step was equal (R~0.88). It means that both models have proper accuracy for predicting CO concentration. However, it is noteworthy that FS-SVM model charged fewer amounts of computational and economical costs due to fewer inputs than SVM model.

Conclusion: Results showed that although both models have relatively equal accuracy in predicting CO concentration, FS-SVM model is the superior model due to its less number of inputs and therefore, less computational burden.

Keywords: Carbon Monoxide, Support Vector Machine, Forward Selection, Tehran

*Corresponding Author:*roohollahnoori@gmail.com*
Tel: +98 9374320526 ***Fax***: +98 21 66407719