

ارزیابی مدل‌های پارامتری و غیر پارامتری در پیش‌بینی وقایع نادر ترافیکی بر مبنای سرعت متوسط و حجم ترافیک

آرش رساءیزدی^۱، سیداحسان سیدابریشمی^{۲*}

- ۱- دکتری برنامه‌ریزی حمل‌ونقل، دانشکده مهندسی عمران و محیط‌زیست، دانشگاه تربیت مدرس
۲- دانشیار برنامه‌ریزی حمل‌ونقل، دانشکده مهندسی عمران و محیط‌زیست، دانشگاه تربیت مدرس

Seyedabrishami@modares.ac.ir

تاریخ پذیرش ۱۴۰۱/۰۳/۲۷

تاریخ دریافت ۱۴۰۰/۰۹/۲۳

چکیده

پیش‌بینی متغیرهای ترافیکی یکی از ابزارهای کارآمد در مدیریت تقاضای سفر است. با استفاده از این ابزار، متغیرهای ترافیکی پیش‌بینی شده در اختیاران کاربران و گردانندگان سیستم حمل‌ونقل قرار می‌گیرد تا برنامه‌ریزی‌های فردی و سیاست‌گذاری‌های کلی اتخاذ شوند. در این پژوهش دو متغیر ترافیکی سرعت متوسط و حجم ترافیک ساعتی، در جاده برون‌شهری کرج به چالوس به‌عنوان محوری با نوسانات زیاد متغیرهای ترافیکی، پیش‌بینی شده است. از میان مدل‌های متنوع پیش‌بینی کننده، مدل ساریما به‌عنوان یک مدل پارامتری و مدل‌های شبکه عصبی مصنوعی و ماشین بردار پشتیبان به‌عنوان مدل‌های غیرپارامتری استفاده شده‌اند. در فرآیند پیش‌پردازش داده، متغیرهای اثرگذار بر سرعت متوسط و حجم ترافیک استخراج و به‌عنوان متغیرهای پیش‌بینی کننده به مجموعه داده اضافه شده است. همچنین از آنجاکه اطلاع داشتن از مقادیر بیشینه و کمینه سرعت متوسط و حجم ترافیک به‌عنوان وقایع نادر ترافیکی، اهمیت بیشتری به نسبت مقادیر عادی دارد، ارزیابی مدل‌ها با تأکید بر پیش‌بینی وقایع نادر انجام شده است. نتایج نشان می‌دهد، برای داده آزمون، کمترین ریشه میانگین مربعات خطای پیش‌بینی سرعت متوسط و حجم ترافیک به ترتیب با استفاده از مدل‌های شبکه عصبی مصنوعی و ماشین بردار پشتیبان و برابر با ۱۳۹ و سیله نقلیه بر ساعت و ۵ کیلومتر بر ساعت حاصل شده است. کم‌ترین ریشه میانگین مربعات خطای پیش‌بینی سرعت متوسط برای چارک اول و چهارم به‌عنوان مقادیر نادر ترافیکی مقادیر مشاهده شده به ترتیب توسط مدل‌های ماشین بردار پشتیبان و شبکه عصبی مصنوعی به دست آمده است. همچنین چارک اول و چهارم مقادیر مشاهده شده حجم ترافیک با مدل ماشین بردار پشتیبان دقیق‌تر از دو مدل دیگر پیش‌بینی شده‌اند.

واژگان کلیدی: پیش‌بینی متغیرهای ترافیکی، پیش‌بینی وقایع نادر، ساریما، شبکه عصبی مصنوعی، ماشین بردار پشتیبان.

۱ - مقدمه

یکی از اطلاعات مورد نیاز برای تمام مسائل برنامه‌ریزی و به طور خاص برنامه‌ریزی حمل‌ونقل از شرایط آینده است. در برنامه‌ریزی حمل‌ونقل به منظور برقراری یا حفظ تعادل میان عرضه و تقاضای حمل‌ونقل، دانستن تقاضا و عرضه در سال مبنا ضرورت دارد. با توجه به نبود اطلاعات دقیق از آینده، این اطلاعات بیشتر با استفاده از روش‌های پیش‌بینی به دست می‌آید. پیش‌بینی‌ها می‌تواند به منظور برنامه‌ریزی‌های با افق زمانی بلندمدت مانند برنامه‌ریزی چهار مرحله‌ای کلاسیک حمل‌ونقل صورت گیرد [1]. گروه دیگر برنامه‌ریزی‌ها، برنامه‌ریزی‌های کوتاه‌مدت است. در این گروه، علاوه بر اینکه افق پیش‌بینی کوتاه‌مدت است، اهداف آن نیز بر مبنای آینده نزدیک و به صورت زودبازده در نظر گرفته می‌شود. یکی از ابزارها در برنامه‌ریزی کوتاه‌مدت، پیش‌بینی کوتاه‌مدت متغیرهای ترافیکی و اطلاع‌رسانی آن به استفاده‌کنندگان و گردانندگان سیستم حمل‌ونقل است. استفاده‌کنندگان از راه می‌توانند بر اساس اطلاعات پیش‌بینی‌شده برنامه‌ریزی شخصی بهتری برای سفرهای آتی خود داشته باشند. منظور از برنامه‌ریزی شخصی انتخاب مسیر و زمان عزیمت توسط فرد مسافر است. چنانچه مقصد فرد مسافر دارای مسیرهای موازی باشد، می‌تواند از بین مسیرهای موازی انتخاب بهینه داشته باشد. از سوی دیگر در صورت نبود محور موازی، فرد مسافر با انتخاب زمان عزیمت مناسب (خارج از بازه‌های اوج سفر) می‌تواند به بهینه شدن سفر خود کمک کند. همچنین گردانندگان سیستم نیز آمادگی بیشتری برای مواجهه با شرایط بحرانی ترافیک دارند و می‌توانند با اعمال سیستم‌های از پیش تعیین شده، مانع بروز تراکم ترافیک و آثار سو ناشی از آن شوند [2].

بیشتر پیش‌بینی کوتاه‌مدت برای دو متغیر ترافیکی سرعت متوسط و حجم ترافیک انجام می‌شود. این یافته بر اساس جستجوها در پایگاه‌های علمی به دست آمده است. نتایج جست و جوی "پیش‌بینی کوتاه مدت سرعت متوسط و حجم ترافیک" به مراتب بیشتر از سایر متغیرهای ترافیکی است. این دو متغیر بیانگر شرایط عملکردی کمان‌های شبکه حمل‌ونقل هستند. به عنوان سایر متغیرهای ترافیک پیش‌بینی شونده می‌توان به زمان

سفر [3]، تراکم ترافیک [4] و وضعیت ترافیک [5] اشاره کرد. به‌عنوان یک نکته قابل توجه، اطلاع داشتن از شرایط بحرانی ترافیک اهمیت و ارزش بیشتری به نسبت اطلاع داشتن از شرایط طبیعی آن دارد. منظور از این شرایط، شرایطی است که مقادیر کمینه و بیشینه سرعت و حجم ترافیک رخ می‌دهد. مقادیر بیشینه حجم ترافیک و کمینه سرعت ترافیک معادل با ازدحام و تراکم ترافیک و مقادیر کمینه حجم ترافیک و بیشینه سرعت ترافیک معادل با وضعیت ناایمن ترافیک است و گرداننده سیستم نیاز دارد ضمن اطلاع از این شرایط، اقدامات مهار کننده سرعت‌های بیشینه را به منظور افزایش ایمنی راه در دستور کار قرار دهد. از آنجاکه در راه‌های برون‌شهری احتمال رخداد چنین شرایط ترافیکی کمتر از شرایط طبیعی است، با عنوان وقایع نادر نیز شناخته می‌شوند [6]. تأکید این پژوهش بر پیش‌بینی وقایع نادر سرعت متوسط و حجم ترافیک است.

جنبه با اهمیت دیگر در پیش‌بینی متغیرهای ترافیکی انتخاب و استفاده از روش‌های متناسب با داده و مسئله است. به طور کلی روش‌های پیش‌بینی به سه گروه، ساده، پارامتری و غیرپارامتری تقسیم‌بندی می‌شوند [7]. مدل‌های ساده، فاقد پیچیدگی و محاسبات سنگین هستند. در مقابل دقت پیش‌بینی آن‌ها به نسبت دو گروه دیگر کمتر است و بیشتر به‌عنوان مدل‌های مبنا استفاده می‌شوند. میانگین تاریخی یکی از روش‌های ساده است که در آن میانگین مقادیر مشاهده شده در زمان‌های مشابه به‌عنوان پیش‌بینی در نظر گرفته می‌شود [8].

مدل‌های پارامتری، مدل‌هایی مبتنی بر آمار و احتمالات هستند. این مدل‌ها تئوری ریاضی قوی‌ای دارند و با تخمین ضرایب و آماره‌ها، قادر به کشف روابط بین متغیرها هستند و مدل‌های تفسیر شونده به حساب می‌آیند. به عنوان نقطه ضعف اصلی این مدل‌ها می‌توان به سازگاری کم با ویژگی‌های کلان داده، شامل حجم، سرعت و تنوع بالا اشاره کرد. فرض محدودکننده رابطه خطی میان متغیرهای مستقل و وابسته نیز یکی دیگر از نقاط ضعف مدل‌های پارامتری است [9]. از جمله شناخته‌ترین مدل‌های پارامتری می‌توان به مدل میانگین متحرک خود همبسته

واقعی نبوده و از خروجی نرم افزارهای شبیه سازی ترافیک استفاده شده است. مان و همکاران [19] عملکرد مدل های شبکه عصبی مصنوعی و آریما را به صورت منفرد و ترکیبی برای پیش بینی حجم ترافیک بررسی کرده اند و نتیجه می گیرند که مدل ترکیبی بر دو مدل منفرد از حیث دقت پیش بینی برتری دارد. به عنوان کاربرد شبکه عصبی در مسائل مرتبط به ایمنی راه نیز می توان به مطالعه جدان و همکاران [20] برای پیش بینی تصادفات جاده ای و مطالعه یو و همکاران [21] برای پیش بینی مدت زمان تصادفات اشاره کرد. مدل پر کاربرد دیگر مدل ماشین بردار پشتیبان است که دوان [22] با استفاده از آن جریان ترافیک را پیش بینی کرده اند. در این مطالعه با بهینه سازی پارامترهای مدل، دقت پیش بینی بیشتری به نسبت مدل پایه حاصل شده است. ژو و همکاران [23] وضعیت ترافیک را با استفاده از ماشین بردار پشتیبان و کالمن فیلتر پیش بینی کرده اند که ماشین بردار پشتیبان این متغیر ترافیکی کیفی را دقیق تر پیش بینی کرده است. همچنین یائو و همکاران [24] برای مطالعه موردی خود نشان دادند که ماشین بردار پشتیبان سرعت ترافیک کوریدورهای شهری را دقیق تر از شبکه عصبی مصنوعی پیش بینی می کند.

در این مطالعه به منظور پیش بینی سرعت متوسط ساعتی و حجم ترافیک ساعتی از هر دو رویکرد پارامتری و غیر پارامتری استفاده شده است. هر کدام از این دو رویکرد نقاط ضعف و قوت متفاوتی دارند. استفاده از هر دو رویکرد به منظور نشان دادن نقاط ضعف رویکرد دیگر انجام شده است. همچنین استفاده از هر دو رویکرد بر جامعیت پژوهش انجام شده می افزاید. بدین منظور، مدل ساریما به عنوان یک مدل پارامتری و مدل های شبکه عصبی مصنوعی و ماشین بردار پشتیبان به عنوان دو مدل غیر پارامتری انتخاب شده اند. داده ترافیکی مورد استفاده مربوط به جاده برون شهری کرج به چالوس، مقطع پورکان است که در طول یک سال و ۶ ماه جمع آوری شده است. علاوه بر دو متغیر پیش بینی شونده، متغیرهای پیش بینی کننده مرتبط، استخراج و به مجموعه داده اضافه شده است. هر دو متغیر سرعت متوسط و

یکپارچه^۱ (آریما) و انواع مشتقات آن از جمله میانگین متحرک خود همبسته یکپارچه فصلی^۲ (ساریما) اشاره کرد. به عنوان نمونه القمدی و همکاران [10] و ژونگ و همکاران [11] به ترتیب از مدل آریما برای پیش بینی تراکم ترافیک و حجم ترافیک استفاده کرده اند و نشان دادند مدل آریما به نسبت مدل های ساده دقت به مراتب بالاتری دارد. لو و همکاران [12] با استفاده از الگوریتم ژنتیک، مدل ساریما بهبود یافته ای را به منظور پیش بینی حجم ترافیک معرفی کرده اند که در مقایسه با مدل ساریما پایه خطای پیش بینی کمتری دارد. برای پیش بینی سرعت ترافیک، یانگ و همکاران [13] مدل های پارامتری و غیر پارامتری را مقایسه و ارزیابی کرده اند و در نهایت به این نتیجه می رسند که مدل های ترکیبی می توانند نقاط ضعف هر دو رویکرد را پوشش دهد و دقت پیش بینی را افزایش دهد. مدل های پارامتری علاوه بر پیش بینی سرعت و حجم برای سایر کاربردها نیز استفاده شده اند. به عنوان نمونه لی و همکاران [14] برای پیش بینی جریان مسافر فرودگاه، میلنکوویچ و همکاران [15] برای پیش بینی جریان مسافر راه آهن و مقیمی و همکاران [16] برای پیش بینی طول چرخه چراغ های راهنمایی و رانندگی از مدل های پارامتری استفاده کرده اند.

مدل های غیر پارامتری بیشتر مبتنی بر یادگیری ماشین هستند. این مدل ها نیز دارای پارامتر هستند اما پارامترها دارای تفسیر و معناداری مشخصی نیستند و به همین دلیل از عبارت مدل های غیر پارامتری استفاده می شود. مدل های غیر پارامتری مدل هایی سازگار با کلان داده ها بود که قادرند هرگونه روابط خطی و غیر خطی را به تصویر بکشند. در مقابل عدم امکان تفسیر روابط علت و معلولی محدودیت اصلی این مدل ها به شمار می رود [17]. مدل های شبکه عصبی مصنوعی^۳، ماشین بردار پشتیبان^۴، نزدیک ترین همسایه^۵ و درخت تصمیم^۶ از جمله شناخته ترین مدل های غیر پارامتری هستند. به عنوان نمونه چسپیکس و همکاران [18] با استفاده از مدل شبکه عصبی مصنوعی جریان ترافیک را پیش بینی کرده اند. داده مورد استفاده در این مطالعه

۴. Support vector machine (SVM)

۵. Nearest neighbours

۶. Decision tree

۱. Autoregressive integrated moving average

۲. Seasonal autoregressive integrated moving average

۳. Artificial neural network (ANN)

اند در حالی که ماهیت متغیرهای پیش‌بینی کننده درون شهری کاملاً متفاوت است. همین نکته سبب اضافه شدن متغیرهای مبتنی بر تعطیلات مانند تعطیل بودن آن روز، نوع تعطیلی، تعطیلی و نوع آن در روزهای قبل و بعد و ایام خاص شمسی و قمری شده است. همچنین در تعریف متغیرهای مربوط به تاریخ از آنجاکه تقویم‌های شمسی و قمری بر هم منطبق نبوده و هر دوی آن بر متغیرهای ترافیکی اثرگذار هستند، هر دو دسته متغیر اضافه شده است. دسته دیگر متغیرهای اضافه شده مربوط به انسداد مسیر در هر دو جهت و انسداد محورهای موازی است. این انسداد بیشتر از طریق محدودیت‌های ترافیکی اعمال شده توسط پلیس راهور اتفاق می‌افتد و به صورت مستقیم بر شرایط ترافیک اثر می‌گذارد. همچنین وضعیت آب‌وهوا به عنوان یک عامل اثرگذار دیگر به مجموعه متغیرها اضافه شده است. جدول (۱) متغیرهای موجود در مجموعه داده را نشان می‌دهد.

جدول ۱. تعریف کلی متغیرهای موجود در مجموعه داده

Feature Name	Description
Season	Including spring, summer, fall, and winter
Solar month	Including 12 solar months
Lunar month	Including 12 lunar months
Day of a solar month	Including 29-31 days of a solar month
Day of a lunar month	Including 29-30 days of a lunar month
Time of day	Including 24 hours a day
6 hours before holidays	Equal to 1 if it is 1 to 6 hours before holidays
6 hours after holidays	Equal to 1 if it is 1 to 6 hours after holidays
Day or night	Including day and night
Number of holidays	The number of sequential holidays
Holidays	Includes 1 for holidays and 0 for other days
Holiday type	Type of holidays
Holiday in three days later	Equal to 1 if three days later is a holiday
Type of holidays in three days later	Including the holiday type of three days later if it is a holiday, otherwise equals 0.
Holiday in three days ago	Equal to 1 if three days ago is a holiday
Type of holidays in three days ago	Including the holiday type of three days ago if it is a holiday, otherwise equals 0.
Holiday in two days later	Equal to 1 if two days later is a holiday
Type of holidays in two days later	Including the holiday type of two days later if it is a holiday, otherwise equals 0.

حجم ترافیک برای یک مقطع ثابت از این محور برداشت شده‌اند و مربوط به کل مسیر نیستند. همچنین علاوه بر ارزیابی کلی دقت مدل‌ها، به شکل خاص دقت پیش‌بینی وقایع نادر نیز بررسی شده است. به‌طورکلی نوآوری این مطالعه شامل موارد زیر است:

- استخراج متغیرهای اثرگذار بر سرعت متوسط و حجم ترافیک و تشکیل بانک اطلاعاتی منسجم و کامل
- به‌کارگیری روش‌های پارامتری و غیر پارامتری و مقایسه عملکرد رویکردها در پیش‌بینی وقایع نادر

۲- معرفی داده

داده مورد استفاده در این مطالعه در مقطع پورکان جاده کرج به چالوس در بازه‌های ساعتی جمع‌آوری شده است. سفرهای انجام شده در این جاده برون‌شهری، واقع در شمال ایران، بیشتر از نوع تفریحی و به سمت مقاصد شمال کشور هستند. سفرهای تفریحی در مقایسه با سفرهای اجباری با اهدافی مانند کار و تحصیل، الگوی پیچیده‌تری دارند و همین نکته بر چالش پیش‌بینی متغیرهای ترافیک آن می‌افزاید. داده خام اولیه تنها شامل، سرعت متوسط، حجم ترافیک، تاریخ و ساعت برداشت اطلاعات است. این داده‌ها توسط سازمان راهداری و حمل‌ونقل جاده‌ای با استفاده از شناسگرهای القایی جمع‌آوری شده و در اختیار نویسندگان قرار گرفته است. پس از بررسی عوامل مؤثر بر متغیرهای ترافیکی در این محور، مشخص شد ارتباط تنگاتنگی میان متغیرهای ترافیکی و تعطیلات وجود دارد. میانگین متغیرهای ترافیکی سرعت متوسط و حجم ترافیک در روزهای تعطیل سال به ترتیب ۲۶ درصد بیشتر و ۱۹ درصد کمتر از میانگین این دو متغیر در روزهای کاری سال هستند. بسیاری از سفرهای برون شهری از جنس سفرهای اختیاری با اهداف غیرکاری هستند. این گونه از سفرهای غیرکاری سفرهای غیر روزمره به حساب می‌آیند و به دنبال آن وابسته به شرایط غیر روزمره هستند. به عبارت دیگر، برای انجام یک سفر تفریحی، فرد مسافر به دنبال یافتن زمان فراغتی است که بیشتر در تعطیلات رسمی رخ می‌دهد. بررسی نوسانات متغیرهای ترافیکی نیز گواهی بر این ادعا است. مطالعات پیشین نیز بیشتر از اطلاعات تقویمی و رخدادها برای پیش‌بینی‌های برون شهری استفاده کرده

Table 2. Statistical description of train and test datasets

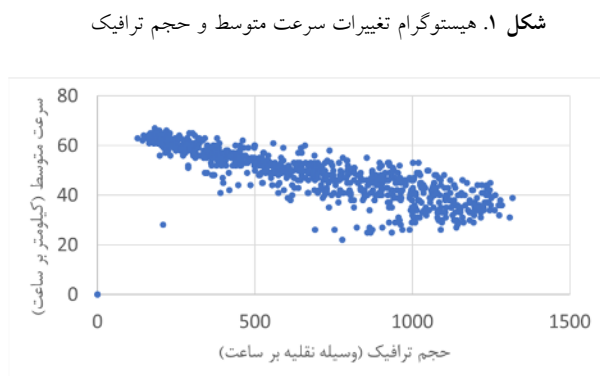


Fig. 1. Histogram of average speed and traffic volume changes

۳- روش‌شناسی

در این بخش ابتدا روش‌شناسی مربوط به سه مدل ساریما، شبکه عصبی مصنوعی و ماشین بردار پشتیبان ارائه می‌شود. در بخش انتهایی این بخش شاخص‌های مورد استفاده به منظور ارزیابی نتایج بررسی می‌شوند.

۳-۱- مدل ساریما

در مدل خود همبسته که به مدل آر^۲ (AR_p) معروف است، مقدار آینده یک متغیر به صورت ترکیب خطی از مشاهدات گذشته (p مشاهده گذشته)، بخش خطای تصادفی و مقدار ثابت مدل در نظر گرفته می‌شود. برخلاف مدل آر که بر اساس مقادیر گذشته سری زمانی تخمین زده می‌شود، مدل میانگین متحرک یا مدل MA_q از خطاهای گذشته به عنوان متغیرهای مستقل (پیش‌بینی کننده) استفاده می‌نماید. مدل آرما^۴ ($ARMA_{pq}$) ترکیبی از مدل‌های آر و ما بوده و برای مدل‌های سری زمانی تک متغیره مفید است [25].

مدل‌های آر، ما و آرما برای سری‌های زمانی ایستا کاربرد داشته درحالی‌که در عمل بیشتر سری‌های زمانی مانند آن‌هایی که به اقتصاد، اجتماع و تجارت مرتبط می‌شوند رفتار نایستا را نشان می‌دهند. یک سری زمانی ایستا است چنانچه میانگین، واریانس و ضریب خودهمبستگی آن در طول زمان ثابت باقی بماند. سری زمانی‌هایی که شامل الگویی روندی یا فصلی هستند، ماهیتی

Holiday in two days ago	Equal to 1 if two days ago is a holiday
Type of holidays in two days ago	Including the holiday type of two days ago if it is a holiday, otherwise equals 0.
Holiday in a day later	Equal to 1 if a day later is a holiday
Type of holidays in a day later	Including the holiday type of a day later if it is a holiday, otherwise equals 0.
Holiday in a day ago	Equal to 1 if a day ago is a holiday
Type of holidays in a day ago	Including the holiday type of a day ago if it is a holiday, otherwise equals 0.
Weather condition	Including sunny, rainy, and snowy
Blockage	Blockage of the road by police
Blockage of the opposite direction	Blockage of the opposite direction by police

Table 1. General description of variables in the database

به جز متغیرهای پیش‌بینی شونده و متغیر تعداد روزهای تعطیل، سایر متغیرها به صورت ساختگی (دامی^۱) تعریف شده‌اند تا امکان ساخت مدل ساریما با استفاده از متغیرهای اسمی فراهم شود. همین موضوع سبب افزایش محسوس تعداد متغیرها شده و برای پرداخت مدل‌ها به پردازنده‌های قوی نیاز پیدا می‌شود.

داده جمع‌آوری شده در طول سال ۱۳۹۷ به عنوان داده آموزش مدل‌ها و داده جمع‌آوری شده در ۶ ماه نخست سال ۱۳۹۸ به عنوان داده آزمون مدل‌ها استفاده شده است. از داده آزمون در فرایند آموزش مدل‌ها استفاده نشده است. داده مورد استفاده مربوط به بازه زمانی پیش از شیوع همه‌گیری ویروس کوید-۱۹ است و در بازه انتخاب شده، این عامل اثری بر متغیرهای ترافیکی ندارد. جدول (۲) تعداد مشاهدات، میانگین و انحراف معیار متغیرهای ترافیکی را به تفکیک داده آزمون و آموزش نشان می‌دهد. همچنین شکل (۱) هیستوگرام تغییرات سرعت متوسط و حجم ترافیک را برای کل داده مورد استفاده را نشان می‌دهد.

جدول ۲. اطلاعات آماری مجموعه داده‌های آموزش و آزمون

Data set	Number of observations	Traffic volume		Average traffic speed	
		Mean	Standard deviation	Mean	Standard deviation
Train	8760	638.80	360.76	49.23	9.23
Test	4380	704.77	357.24	47.67	10.06

۳. Moving average

۴. Autoregressive moving average

۱. Dummy

۲. Autoregressive

نرمالیزه کردن از جمله رایج‌ترین روش پیش‌پردازش روی داده‌ها است که سبب کاهش میزان تفاوت در داده‌ها و موجب بهبود فرآیند یادگیری می‌شود. در این روش داده‌ها چنان تغییر می‌یابند که دامنه تغییرات داده‌ها در محدوده $[L, H]$ باشد. این کار با استفاده از رابطه ۳ انجام می‌گیرد [26].

$$X_n = \frac{x_i - X_{min}}{X_{max} - X_{min}} \times (H - L) + L \quad (3)$$

$$i = 1, 2, \dots, n$$

در این رابطه X_i مقدار واقعی ورودی شبکه و X_n مقدار نرمال شده متناظر با آن است. X_{min} و X_{max} به ترتیب مقدار کمینه و بیشینه هستند. در این پژوهش از دامنه $[-1, 1]$ به‌عنوان محدوده نرمال‌سازی استفاده شده است.

۳-۳-۳ مدل ماشین بردار پشتیبان

ماشین بردار پشتیبان یکی از الگوریتم‌های شناخته شده در زمینه پیش‌بینی متغیرهای ترافیکی است. بردار پشتیبان در واقع دسته‌بندی است که مرز دسته‌ها را مشخص می‌کند. در مدل ماشین بردار پشتیبان با معیار قرار دادن داده‌های قرار گرفته در مرز به عنوان بردارهای پشتیبان، به دنبال یافتن بهترین صفحات دسته‌بندی است به شکلی که فاصله بردارهای پشتیبان تا صفحات پشتیبان در فضای چندبعدی بیشینه شود. همچنین این الگوریتم به سایر نقاط قرار گرفته در جایی غیر از مرز دسته‌ها حساس نیست [28].

در حالت پیچیده‌تر که داده‌ها به صورت غیرخطی پراکنده شده باشند لازم است داده‌ها به کمک تابع ریاضی کرنل^۲ به فضای تفکیک‌پذیر دیگر منتقل شوند تا پراکندگی مشاهدات به صورت خطی درآیند به‌گونه‌ای که یافتن صفحات پشتیبان در فضای چندبعدی، آسان شود. در این مطالعه از تابع کرنل پایه شعاعی^۳ (RBF) مطابق با رابطه ۴ استفاده شده است [28].

$$K(X_i, X_j) = \exp\left(-\frac{\|X_i - X_j\|^2}{2\gamma^2}\right) \quad (4)$$

که در آن γ (گاما) پارامتر آزاد مدل است و $\|X_i - X_j\|^2$ مجذور فاصله اقلیدسی میان دو بردار ویژگی X_i و X_j است.

نایستا دارند. به همین منظور مدل آریمما پیشنهاد شده که تعمیم یافته مدل آرما بوده و شامل موارد نایستا می‌شود. در مدل آریمما، سری‌های زمانی نایستا با d بار تفاضل‌گیری، ایستا می‌شود [25].

مدل آریمما برای داده‌های غیرفصلی کاربرد دارد. باکس و جنکینس این مدل را گسترش داده تا به صورت فصلی استفاده شود. مدل پیشنهادی ساریما نامیده شده است. هرگاه در یک سری بعد از فاصله زمانی مشخص (S)، شباهت‌هایی پیدا شود، سری دارای رفتار فصلی یا تناوبی با دوره تناوب S می‌شود. در این مدل تفاوت‌های فصلی با درجه مناسب استفاده می‌شود تا نایستا بودن سری زمانی مرتفع شود. تفاوت درجه فصلی، اختلاف بین یک مشاهده و مشاهده منطبق آن در سال گذشته است [25].

۳-۲-۲ مدل شبکه عصبی مصنوعی

مدل‌های شبکه عصبی مصنوعی یکی از روش‌های پیش‌بینی متغیرهای است که هدف اساسی آن ساخت یک مدل با الگوگیری از عملکرد مغز بشر است. یکی از پرکاربردترین شبکه عصبی مصنوعی برای پیش‌بینی، پرسپترون چندلایه است که از یک لایه ورودی، یک یا چند لایه نهان و یک لایه خروجی برای پیش بردن شبکه استفاده می‌کند. نرون‌ها (گره‌ها) در لایه‌های مختلف به عنوان المان‌های پردازشی شناخته می‌شوند. هر نرون در لایه‌های نهان، خروجی نرون‌های متصل شده در لایه قبلی را گرفته و با جمع کردن وزن‌دار (رابطه ۱) ورودی‌های گرفته شده آن را به عنوان خروجی به لایه بعد انتقال می‌دهد [26].

$$I_{j(k)} = \sum_{i=1}^M w_{ij} \sigma w_{i(k-1)} O_{i(k-1)} + \gamma \quad (1)$$

که در آن $I_{j(k)}$ و $O_{j(k)}$ به ترتیب ورودی و خروجی نرون j -ام در لایه نهان k -ام است. w_{ij} بردار وزن خروجی نرون i به عنوان ورودی نرون j است. γ بخش خطا و $\sigma()$ تابع تبدیل سیگموئید^۱ است که مطابق رابطه ۲ تعریف می‌شود [27].

$$\sigma(X) = \frac{1}{1 + \exp(-X)} \quad (2)$$

۳. Radial Basis Function

۱. Sigmoid transform function

۲. Kernel function

۳-۴- ارزیابی دقت مدل‌ها

به منظور ارزیابی دقت روش‌های پیش‌بینی سرعت و حجم ترافیک به عنوان دو متغیر پیوسته، شاخص‌های متنوعی وجود دارد. شاخص‌های زیر از جمله مرسوم‌ترین این شاخص‌ها به حساب می‌آیند. ریشه میانگین مربعات خطا (RMSE): این شاخص بیانگر ریشه دوم متوسط مربعات خطا بین مشاهده و پیش‌بینی است (رابطه ۵). چنانچه ریشه دوم گرفته نشود میانگین مربعات خطا (MSE) به دست می‌آید [29].

$$RMSE = \sqrt{\frac{\sum_{t=1}^n (\hat{y}_t - y_t)^2}{n}} \quad (5)$$

که در آن، y_t ، \hat{y}_t و n به ترتیب مقادیر مشاهده شده، پیش‌بینی شده و تعداد کل مشاهدات هستند.

ضریب تعیین نمودار مشاهده-برآورد: برای محاسبه این شاخص، ابتدا موقعیت مشاهدات و مقادیر پیش‌بینی شده در دستگاه مختصات دو بعدی مشخص می‌شود. سپس معادله خط رگرسیون عبوری از این نقاط پیدا شده و با استفاده از رابطه ۶، ضریب تعیین نمودار مشاهده-برآورد محاسبه می‌شود [30].

$$R^2 = 1 - \frac{ESS}{TSS} \quad (6)$$

که در آن ESS مجموع مربعات خطا^۲ و TSS کل مجموع مربعات^۴ است.

۴- نتایج

در گام نخست مدل‌سازی تعیین مقادیر بهینه پارامترهای مدل‌ها ضرورت دارد. مقادیر بهینه با بررسی مقادیر مختلف و ارزیابی دقت حاصله برای داده آزمون انتخاب می‌شوند. در مدل ماشین بردار پشتیبان دو پارامتر موجود، گاما و هزینه (C) در تابع بهینه-ساز است که پس از بررسی مقادیر مختلف به ترتیب مقادیر ۱ و ۳۲ انتخاب شده است. تعداد لایه‌های بهینه در مدل شبکه عصبی مصنوعی ۴ لایه است. انتخاب تعداد لایه‌ها با سعی و خطا و با هدف کمینه کردن خطای پیش‌بینی انجام شده است. تعداد لایه‌ها از ۱ تا ۲۰ افزایش یافته و در نهایت ۴ لایه کمترین خطای

پیش‌بینی را برای داده‌های آزمون و آموزش را به همراه داشته است. همچنین مدل نهایی ساریما به صورت، $Arima(1,0,0)$ و $Sarima(1,0,1)$ استفاده شده است. یک سری زمانی ایستا به طور تصادفی حول یک میانگین ثابت نوسان می‌کند ولی در سری زمانی پویا هیچ میانگین ثابتی وجود ندارد. در این پژوهش به این منظور از آزمون دیکی-فولر تعمیم‌یافته^۵ برای تعیین ایستایی مدل استفاده شده است [31]. مقادیر آماره t در آزمون دیکی فولر تعمیم‌یافته در جدول (۳) نشان داده شده است.

جدول ۳. آماره t در آزمون دیکی فولر برای تعیین ایستایی مدل پیش‌بینی حجم و سرعت ترافیک

Traffic parameter	Traffic volume	Average traffic speed
Augmented Dickey-Fuller test	-23.63	-22.64
Level 1%	-3.43	
Level 5%	-2.86	
Level 10%	-2.57	

Table 3. T statistic in the Dickey Fuller test to determine the stationarity of the traffic parameters

همان‌طور که در جدول (۳) مشخص است، سری زمانی برای هر دو متغیر ترافیکی ایستا است، زیرا آماره t دیکی-فولر از مقادیر بحرانی مربوط به سطوح معنی‌دار کم‌تر است. جدول (۴) نتایج پیش‌بینی سرعت متوسط و حجم ترافیک به تفکیک داده‌های آزمون و آموزش را نشان می‌دهد.

جدول ۴. شاخص‌های خطای پیش‌بینی به ازای مدل‌های پیش‌بینی کننده متغیرهای ترافیکی

Model	Traffic parameter	Data set	R ²	RMSE
SARIMA	Volume	Train	0.839	143
		Test	0.811	158
	Speed	Train	0.767	6
		Test	0.712	8
ANN	Volume	Train	0.843	107
		Test	0.820	192
	Speed	Train	0.782	4
		Test	0.753	5
SVM	Volume	Train	0.937	88
		Test	0.837	139
	Speed	Train	0.876	5
		Test	0.769	7

۴. Total sum of square

۵. Augmented Dickey-Fuller

۱. Root Mean Square Error

۲. Mean Square Error

۳. Error sum of square

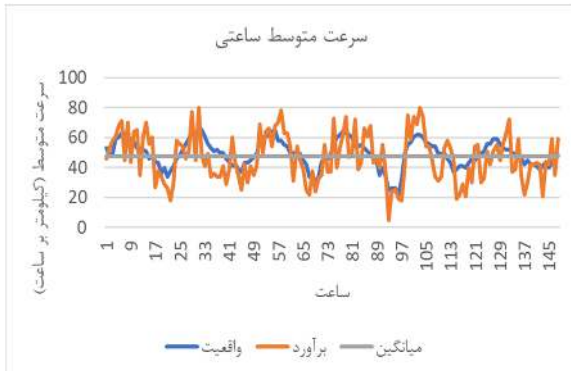


Fig. 2. Changes in speed and volume of observed traffic, estimated based on the least error model and average observations

از آنجا که پيش بينی مقادير بیشینه و کمینه سرعت متوسط و حجم ترافیک از اهمیت بیشتری برخوردار است، ریشه مربعات خطا (RSE) به ازای پيش بينی‌های حجم و سرعت محاسبه شده است. در نمودارهای ارائه شده در شکل (۳ و ۴)، محور افقی مقادير حجم ترافیک و سرعت متوسط و محور عمودی ریشه مربعات خطا است. در این نمودارها نواحی با رنگ روشن‌تر نشان‌دهنده تراکم بیشتر داده‌ها است. هرچه نواحی روشن‌تر به سمت پایین نمودار متمرکز شوند تعداد مشاهدات با مقدار ریشه مربعات خطای کم، بیشتر می‌شود. در مقابل نقاط تاریک تر بدین معنا است که تعداد مشاهدات با مقدار ریشه مربعات خطای کم، کمتر است [34-32]

شکل ۳. تغییرات ریشه مربعات خطا به ازای مقادير مختلف حجم ترافیک به تفکیک مدل‌ها

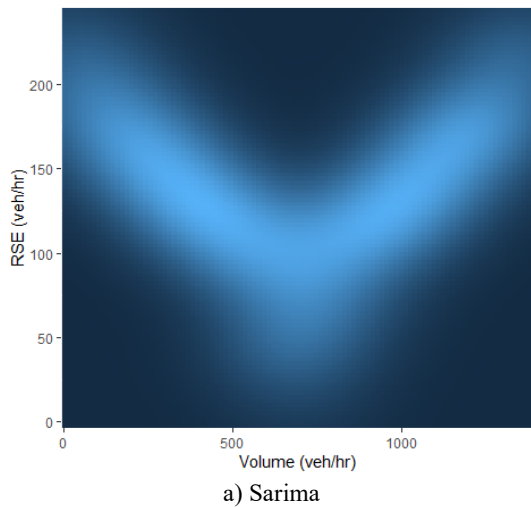
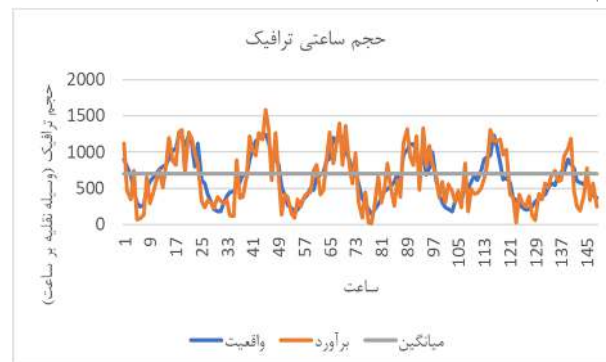


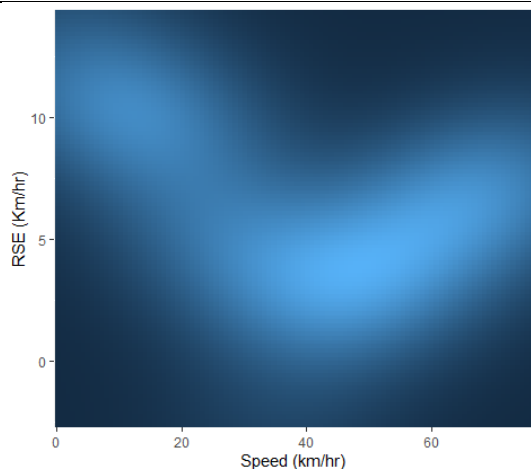
Table 4. Prediction error indices for predicting traffic variables

بر اساس معيار ضريب تعيين، دقت پيش بينی سرعت متوسط و حجم ترافیک با مدل ساراما برای داده آزمون یکسان است. در مقابل ضريب تعيين نمودار برآورد-مشاهده مدل حجم ساعتی ترافیک برای داده آموزش، بیشتر است. مدل شبکه عصبی مصنوعی و ماشین بردار پشتیبان، بر اساس شاخص ضريب تعيين، حجم ترافیک را دقیق‌تر از سرعت متوسط ترافیک پيش بينی کرده‌اند. بر اساس شاخص ریشه میانگین مربع خطا کمترین خطای پيش بينی سرعت متوسط و حجم ترافیک برای داده آزمون به ترتیب با استفاده از شبکه عصبی مصنوعی و ماشین بردار پشتیبان به دست آمده است. به طور کل بر اساس شاخص‌های ارائه شده در جدول (۴) عملکرد دو مدل غير پارامتری از عملکرد مدل پارامتری ساراما بهتر است.

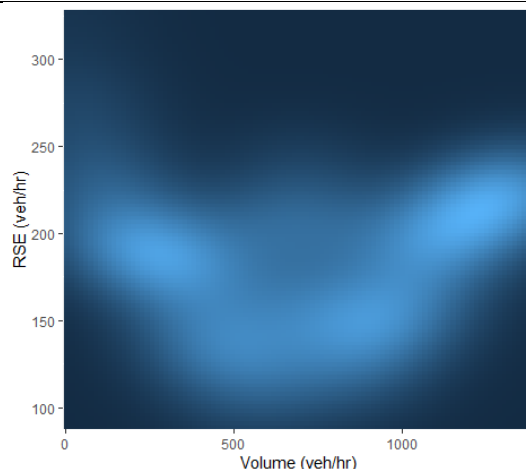
شکل (۲) تغییرات سرعت و حجم ترافیک مشاهده شده، برآورد شده بر اساس کم خطاترین مدل (در پاراگراف قبل معرفی شده است) و میانگین مشاهدات را برای مشاهدات هفته اول داده آزمون را نشان می‌دهد. دلیل انتخاب بازه زمانی کوچک برای این دو شکل وضوح بیشتر تغییرات در نمودارها است. این شکل نشان می‌دهد مدل‌های پيش بينی کننده به مراتب دقت بیشتری به نسبت میانگین مشاهدات دارند.

شکل ۲. تغییرات سرعت و حجم ترافیک مشاهده شده، برآورد شده بر اساس کم خطاترین مدل و میانگین مشاهدات

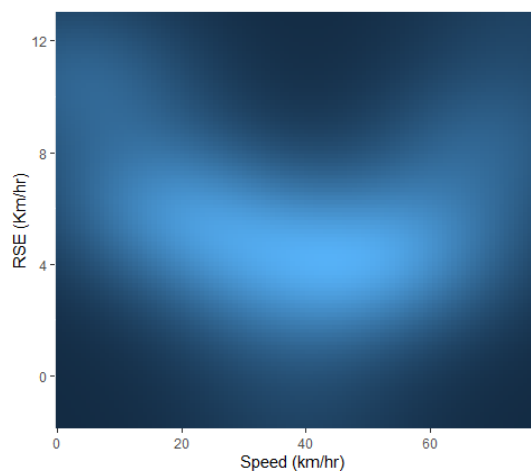




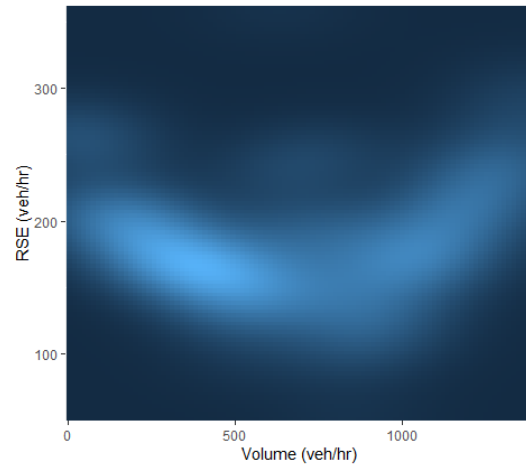
b) ANN



b) ANN



c) SVM



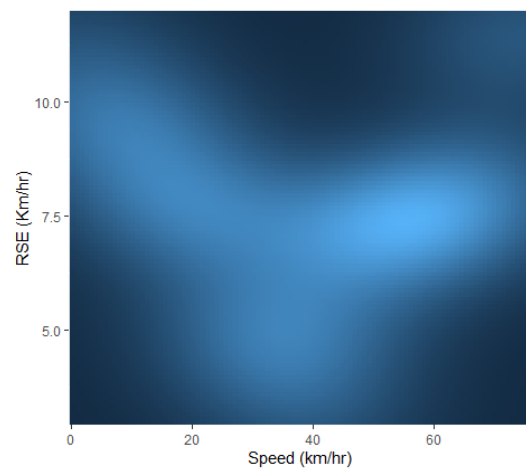
c) SVM

Fig. 4. The changes in the RSE for different values of the average speed for each model

با توجه به شکل (۳ و ۴)، در مدل ساریما، برای هر دو مدل سرعت متوسط و حجم ترافیک، مدل به ازای مقادیر بیشینه و کمینه خطای پیش‌بینی بیشتری دارد. اغلب پیش‌بینی‌های انجام شده برای سرعت مشاهده شده حدود ۴۰ کیلومتر بر ساعت ریشه مربعات خطایی حدود ۴ کیلومتر بر ساعت دارند و این مقدار برای مقادیر ابتدایی و انتهایی سرعت مشاهده شده به ترتیب به حدود ۱۰ و ۷/۵ کیلومتر بر ساعت می‌رسد. به عبارت دیگر مدل ساریما مقادیر بیشینه سرعت را با خطای کمتری به نسبت مقادیر کمینه پیش‌بینی می‌کند. در پیش‌بینی حجم ترافیک وضعیت مشابهی برای پیش‌بینی مقادیر کمینه و بیشینه وجود دارد و ریشه مربعات خطای پرتکرار در این دو محدوده بین ۱۵۰ تا ۲۰۰ وسیله نقلیه بر ساعت است درحالی‌که برای حجم ترافیک

Fig. 3. The changes in the RSE for different values of the traffic volume for each model

شکل ۴. تغییرات ریشه مربعات خطا به ازای مقادیر مختلف سرعت متوسط ترافیک به تفکیک مدل‌ها



a) SARIMA

جدول ۵. خطای پیش‌بینی سرعت متوسط و حجم ترافیک برای چارک اول و چهارم مشاهدات

Model	RMSE Speed		RMSE Volume	
	First quartile	Forth quartile	First quartile	Forth quartile
SARIMA	8.5	8.4	174	166
ANN	9.8	6.1	202	220
SVM	7.7	7.5	169	162

Table 5. Average speed and traffic volume prediction error for the first and fourth quarter of observations

جدول (۵) نشان می‌دهد کم‌ترین ریشه میانگین مربعات خطا پیش‌بینی سرعت متوسط برای چارک اول و چهارم به ترتیب توسط مدل‌های ماشین بردار پشتیبان و شبکه عصبی مصنوعی به دست آمده است. به عبارت دیگر مدل ماشین بردار پشتیبان سرعت‌های کمینه را دقیق‌تر از دو مدل دیگر پیش‌بینی می‌کند و سرعت‌های بیشینه با دقت بیشتری توسط مدل شبکه عصبی مصنوعی پیش‌بینی می‌شود. در پیش‌بینی حجم ترافیک هر دو گروه مشاهدات بیشینه و کمینه حجم ترافیک با مدل ماشین بردار پشتیبان دقیق‌تر از دو مدل دیگر پیش‌بینی شده‌اند.

۵ - نتیجه‌گیری

مسئله پیش‌بینی و برنامه‌ریزی کوتاه‌مدت یکی از مسائل مورد توجه پژوهشگران حمل‌ونقل است. به‌طور کلی با داشتن اطلاعات شرایط فعلی و پیش‌بینی شرایط آینده می‌توان قبل از وقوع هر رخداد در خصوص آن چاره‌اندیشی کرد. زمانی که پارامترهای ترافیکی آینده پیش‌بینی شود، می‌توان برای رفع مشکل تراکم ترافیک هم از طریق اطلاع‌رسانی به استفاده‌کنندگان و هم از طریق اعمال طرح‌های موقت از سوی تصمیم‌گیرندگان سیستم، مانند یک‌طرفه کردن مسیر به مقابله با مشکل قریب‌الوقوع پرداخت.

در این پژوهش به منظور پیش‌بینی کوتاه‌مدت سرعت متوسط و حجم ترافیک در محور برون‌شهری چالوس به کرج، ابتدا پس از دریافت خام اولیه، عوامل اثرگذار بر این دو متغیر ترافیکی شناسایی و به مجموعه داده اضافه شد. این عوامل بیشتر بر مبنای زمان، تاریخ، ایام خاص و تعطیلات شمسی و قمری، انسدادها و آب‌وهوا تعریف شده‌اند. در مرحله بعد متغیرهای

حدود ۷۵۰ وسیله نقلیه بر ساعت این مقدار به ۵۰ وسیله نقلیه در ساعت می‌رسد.

در مدل شبکه عصبی مصنوعی کمترین خطای پیش‌بینی حجم برای مشاهدات واقعی در بازه ۳۰۰ تا ۸۰۰ وسیله نقلیه رخ داده و برابر با ۱۰۰ وسیله نقلیه بر ساعت است. در حد بالای حجم مشاهده شده ریشه مربعات خطای پیش‌بینی پرتکرار حدود ۲۳۰ وسیله نقلیه و در حد پایین این مقدار برابر با ۲۰۰ وسیله نقلیه بر ساعت است. در شکل (۴)، تمرکز ریشه مربعات خطا در حد پایین بالاتر از حد بالای سرعت متوسط است و مدل شبکه عصبی سرعت‌های بیشینه را با خطای کمتری به نسبت سرعت‌های کمینه پیش‌بینی می‌کند. همچنین در حد وسط سرعت مشاهده شده ریشه مربعات خطا در حدود ۳ کیلومتر بر ساعت است.

مدل ماشین بردار پشتیبان عملکرد نسبتاً ثابتی به ازای تمام مقادیر حجم ترافیک دارد. ریشه مربعات خطا پرتکرار به ازای احجام ترافیک مختلف در محدوده ۱۵۰ تا ۲۲۰ وسیله نقلیه بر ساعت قرار می‌گیرد. با این حال خطای پیش‌بینی پرتکرار در نقاط ابتدایی و انتهایی محور حجم ترافیک بیشتر از نقاط میانی آن است. در مقابل در پیش‌بینی سرعت متوسط، پراکندگی نقاط بیشتر بوده و پرتکرارترین ریشه مربعات خطا در محدوده ۲ تا ۱۲ کیلومتر بر ساعت به ترتیب به ازای سرعت‌های میانی و سرعت‌های بحرانی تغییر می‌کند.

به منظور ارزیابی و مقایسه مدل‌ها در خصوص پیش‌بینی سرعت و حجم کمینه و بیشینه، ریشه میانگین مربعات خطا به ازای چارک اول و چهارم مقادیر مشاهده‌شده سرعت متوسط و حجم ترافیک محاسبه و در جدول (۵) ارائه شده است. مقادیر چارک اول و چهارم مشاهدات معادل مقادیر بیشینه و کمینه سرعت و حجم و به تبع آن تراکم ترافیک، خلوتی مسیر، سرعت‌های بالا و نایمن و زمان سفر زیاد هستند پس به عنوان وقایع نادر بررسی آن‌ها ضرورت دارد.

3. Kwak S, Geroliminis N. Travel time prediction for congested freeways with a dynamic linear model. *IEEE Transactions on Intelligent Transportation Systems*. 2020.
4. Akhtar M, Moridpour S. A Review of Traffic Congestion Prediction Using Artificial Intelligence. *Journal of Advanced Transportation*. 2021;2021.
5. Li W, Wang J, Fan R, Zhang Y, Guo Q, Siddique C, et al. Short-term traffic state prediction from latent structures: Accuracy vs. efficiency. *Transportation Research Part C: Emerging Technologies*. 2020;111:72-90.
6. Befekadu GK. Rare Event Simulation in a Dynamical Model Describing the Spread of Traffic Congestions in Urban Network Systems. *A Collection of Papers on Chaos Theory and Its Applications: IntechOpen*; 2021.
7. Lefèvre S, Sun C, Bajcsy R, Laugier C, editors. Comparison of parametric and non-parametric approaches for vehicle speed prediction. 2014 American Control Conference; 2014: IEEE.
8. Campbell JY, Thompson SB. Predicting excess stock returns out of sample: Can anything beat the historical average? *The Review of Financial Studies*. 2008;21(4):1509-31.
9. Zhang Y, Liu Y. Comparison of parametric and nonparametric techniques for non-peak traffic forecasting. *World Academy of Science, Engineering and Technology*. 2009;51(27):8-14.
10. Alghamdi T, Elgazzar K, Bayoumi M, Sharaf T, Shah S, editors. Forecasting traffic congestion using ARIMA modeling. 2019 15th International Wireless Communications & Mobile Computing Conference (IWCMC); 2019: IEEE.
11. Zhong-Da T, Shu-Jiang L, Yan-Hong W, Xiang-Dong W. Network traffic prediction based on ARIMA with Gaussian process regression compensation. *Journal of Beijing University of Posts and Telecommunications*. 2017;40(6):65.
12. Luo X, Niu L, Zhang S. An algorithm for traffic flow prediction based on improved SARIMA and GA. *KSCE Journal of Civil Engineering*. 2018;22(10):4107-15.
13. Yang X, Zou Y, Tang J, Liang J, Ijaz M. Evaluation of short-term freeway speed prediction based on periodic analysis using statistical models and machine learning models. *Journal of Advanced Transportation*. 2020;2020.
14. Li Z, Bi J, Li Z, editors. Passenger flow forecasting research for airport terminal based on SARIMA time series model. *IOP conference series: earth and environmental science*; 2017: IOP Publishing.

ترافیکی شامل سرعت متوسط حجم ترافیک با استفاده از دو رویکرد پارامتری و غیرپارامتری پیش‌بینی و نتایج مقایسه و ارزیابی شدند. مدل ساریما به عنوان یک مدل پارامتری و مدل‌های شبکه عصبی مصنوعی و ماشین بردار پشتیبان به عنوان مدل‌های غیرپارامتری به کار گرفته شد. همچنین با توجه به اهمیت پیش‌بینی مقادیر بحرانی سرعت متوسط و حجم ترافیک به عنوان وقایع نادر، ارزیابی مدل‌ها با تأکید بر چنین مشاهداتی انجام شد. نتایج نشان می‌دهد کمترین خطای پیش‌بینی کلی سرعت متوسط و حجم ترافیک به ترتیب استفاده از مدل‌های شبکه عصبی مصنوعی و ماشین بردار پشتیبان به دست آمده است. در مقابل بیشترین دقت پیش‌بینی برای چارک اول و چهارم مقادیر مشاهده شده سرعت متوسط ترافیک به ترتیب با استفاده از ماشین بردار پشتیبان و شبکه عصبی مصنوعی و برای چارک اول و چهارم مقادیر مشاهده شده حجم ترافیک با استفاده از مدل ماشین بردار پشتیبان حاصل شده است. به عنوان پیشنهاد برای ادامه این مطالعه استفاده از مدل‌های ترکیبی پارامتری و غیرپارامتری که نقاط ضعف هر رویکرد را پوشش می‌دهند، توصیه می‌شود.

۶- سپاس‌گزاری

بدین وسیله از مسئولین و کارشناسان سازمان راهداری و حمل‌ونقل جاده‌ای بابت حمایت از این پژوهش و در اختیار قرار دادن داده ترافیکی راه‌های برون شهری تشکر و قدردانی می‌شود.

۷- اعلام تعارض منافع

نویسندگان اعلام می‌کنند که هیچ نوع تعارض منافی وجود ندارد.

منابع

1. Ahmed B. The Traditional Four Steps Transportation Modeling Using a Simplified Transport Network: A Case Study of Dhaka City, Bangladesh. *International Journal of Advanced Scientific Engineering and Technological Research*. 2012;1(1):19-40.
2. Nagy AM, Simon V. Survey on traffic prediction in smart cities. *Pervasive and Mobile Computing*. 2018;50:148-63.

- forecasting population density. *IJISTECH (International Journal of Information System & Technology)*. 2017;1(1):43-54.
28. Pisner DA, Schnyer DM. Support vector machine. *Machine Learning: Elsevier*; 2020. p. 101-21.
 29. Alzahrani SI, Aljamaan IA, Al-Fakih EA. Forecasting the spread of the COVID-19 pandemic in Saudi Arabia using ARIMA prediction model under current public health interventions. *Journal of infection and public health*. 2020;13(7):914-9.
 30. Stock JH, Watson MW. *Introduction to econometrics 3rd ed.* Pearson Education; 2015.
 31. Im KS, Pesaran MH, Shin Y. Testing for unit roots in heterogeneous panels. *Journal of econometrics*. 2003;115(1):53-74.
 32. Rasaizadi, A., Ardestani, A., & Seyedabrishami, S. E. (2021). Traffic management via traffic parameters prediction by using machine learning algorithms. *International Journal of Human Capital in Urban Management*, 6(1), 57-68.
 33. Rasaizadi, A., Seyedabrishami, S., & Saniee Abadeh, M. (2021). Short-Term Prediction of Traffic State for a Rural Road Applying Ensemble Learning Process. *Journal of Advanced Transportation*, 2021.
 34. Rasaizadi, A., Sherafat, E., & Seyedabrishami, S. (2021). Short-term prediction of traffic state, statistical approach versus machine learning approach. *Scientia Iranica*.
 15. Milenković M, Švadlenka L, Melichar V, Bojović N, Avramović Z. SARIMA modelling approach for railway passenger flow forecasting. *Transport*. 2018;33(5):1113-20.
 16. Moghimi B, Safikhani A, Kamga C, Hao W. Cycle-length prediction in actuated traffic-signal control using ARIMA model. *Journal of Computing in Civil Engineering*. 2018;32(2):04017083.
 17. Wang J, Boukerche A. Non-parametric models with optimized training strategy for vehicles traffic flow prediction. *Computer Networks*. 2021;187:107791.
 18. Csikós A, Viharos ZJ, Kis KB, Tettamanti T, Varga I, editors. Traffic speed prediction method for urban networks—an ANN approach. 2015 International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS); 2015: IEEE.
 19. Man-Chun T, Luo-Bin F, Jian-Min X. Traffic flow prediction based on hybrid ARIMA and ANN model. *China Journal of Highway and Transport*. 2007;4(86).
 20. Jadaan KS, Al-Fayyad M, Gammoh HF. Prediction of road traffic accidents in Jordan using artificial neural network (ANN). *Journal of Traffic and Logistics Engineering*. 2014;2(2).
 21. Yu B, Wang Y, Yao J, Wang J. A comparison of the performance of ANN and SVM for the prediction of traffic accident duration. *Neural Network World*. 2016;26(3):271.
 22. Duan M, editor Short-time prediction of traffic flow based on PSO optimized SVM. 2018 international conference on intelligent transportation, big data & smart city (ICITBS); 2018: IEEE.
 23. Xu D-w, Wang Y-d, Jia L-m, Qin Y, Dong H-h. Real-time road traffic state prediction based on ARIMA and Kalman filter. *Frontiers of Information Technology & Electronic Engineering*. 2017;18(2):287-302.
 24. Yao B, Chen C, Cao Q, Jin L, Zhang M, Zhu H, et al. Short-term traffic speed prediction for an urban corridor. *Computer-Aided Civil and Infrastructure Engineering*. 2017;32(2):154-69.
 25. Brockwell PJ, Brockwell PJ, Davis RA, Davis RA. *Introduction to time series and forecasting*: Springer; 2016.
 26. Heriz HH, Salah HM, Abdu SBA, El Sbihi MM, Abu-Naser SS. *English Alphabet Prediction Using Artificial Neural Networks*. 2018.
 27. Wanto A, Windarto AP, Hartama D, Parlina I. Use of binary sigmoid function and linear identity in artificial neural networks for

Evaluation of parametric and non-parametric models in predicting rare traffic events based on average speed and traffic volume

Arash Rasaizadi¹, Seyedehsan Seyedabrishami^{2*}

- 1- Ph.D. in Transportation Planning, Tarbiat Modares University, Tehran, Iran
- 2- Associate Professor of Transportation Planning, Tarbiat Modares University, Tehran, Iran

Abstract

One of the information needed for all planning problems and specifically transportation planning is to have accurate prediction about the future. Traffic variables prediction is one of the efficient tools in travel demand management. Using this tool and advanced traveler information systems (ATIS), the predicted traffic variables are informed to users and transportation system operators to make plans and set policies. In this study, the average speed and traffic volume of the Karaj to Chalus suburban road with the high variation of traffic variables in the north of Iran is predicted. The Karaj to Chalous road is part of the route from Tehran as the capital of Iran to the country's northern coast. Along the Karaj to Chalous road, three parallel roads, with different lengths, connect Tehran with the cities of the north. In general, finding the pattern of non-mandatory trips is more complicated than mandatory trips. Generally, the predictive methods are divided into three groups, naïve, parametric and non-parametric methods. Among the various predictive models, the SARIMA as a parametric model and the artificial neural network and the support vector machine as nonparametric models are employed. In the data pre-processing step, the variables affecting the average speed and traffic volume are extracted and added to the dataset as predictor variables. These variables are related to time, calendar, holidays, weather, and roads blockage. Also, because of the importance of the maximum and minimum values of traffic speed and volume, as critical values and rare events, models are evaluated with emphasis on the prediction of rare events compared to normal values. The results show that, for the test data, the lowest root mean square error of predicting the average traffic speed and traffic volume are obtained using artificial neural network and support vector machine models equals 139 vehicles per hour and 5 kilometers per hour, respectively. In terms of R^2 of prediction-observation plot, the performance of SARIMA for predicting the average speed and traffic volume is the same for the test dataset. In contrast the R^2 of hourly traffic volume prediction is higher for the training data. The R^2 of artificial neural network model and the support vector machine for traffic volume prediction is higher than traffic speed prediction. The lowest root mean square error of predicting the first and fourth quartile of the observed average traffic speed values is obtained by support vector machine models and artificial neural network, respectively. Also, predicting the first quartile and fourth quartile of the observed traffic volume values by the support vector machine model is more accurate than two other models. Using predicted traffic parameters and providing them to travelers and transportation agencies by intelligent transportation systems leads to make a balance between travel demand and travel supply in the near future which is the main aim of this study. Travelers can have a better personal plan for their future trips based on these predictions. Also, the transportation agencies are more prepared to deal with critical traffic situations and can prevent traffic congestion.

Keywords: Traffic variables prediction, Rare event prediction, SARIMA, Artificial neural network, Support vector machine.