فصلنامه مهندسی ترافیک، سال بیستم، شماره ۸۱، تابستان ۱۳۹۹، ص. ۴۷-۳۵

توسعه الگوریتم‌های داده‌کاوی جهت تعیین و پیش‌بینی وضعیت ترافیک در مسیرهای برون‌شهری

وحید برادران (مسئول مکاتبات)، استادیار، گروه مهندسی صنایع دانشگاه آزاد اسلامی واحد تهران شمال، تهران، ایران

حمید برادران، کارشناس ارشد مهندسی عمران گرایش برنامه‌ریزی حمل و نقل، شرکت کنترل ترافیک تهران، ایران

E-mail: v_baradaran@iau-tnb.ac.ir

چکیده

تعیین و پیش‌بینی وضعیت ترافیک در یک ساعت به کمک مقدار متغیرهای ترافیکی (جریان، سرعت و چگالی) که در یک مسیر اندازه‌گیری می‌شوند، به برنامه‌ریزان شهری جهت استفاده بهینه امکانات و منابع و به رانندگان در انتخاب مسیرهای مناسب کمک می‌کند. توسعه ابزاری جهت تحلیل تغییرات همزمان دو یا چند متغیر ترافیکی در یک مسیر و توسعه قواعد ثابتی که به کمک آنها بتوان وضعیت ترافیک را جهت اطلاع به رانندگان مشخص کرد، هدف این مقاله است. در این مقاله با استفاده از رویکرد خوشه‌بندی، یکی از تکنیک‌های داده‌کاوی، ساعات مختلف روزهای گذشته بر اساس تشابه مقدار متوسط سرعت و جریان در آنها خوشه‌بندی شده‌اند. تحلیل مقدار متغیرهای ترافیکی در یک خوشه، امکان تفسیر وضعیت ترافیک و استخراج قوانینی برای تعیین وضعیت ترافیک در آینده را در یک مسیر مشخص ساده‌تر می‌کند. بر اساس رویکرد پیشنهادی، ساعات تردد خودروها در یکی از بزرگراه‌های برون‌شهری (ساوه- تهران) بر اساس داده‌های گذشته خوشه‌بندی شده‌اند. بر این اساس، چهار وضعیت ترافیکی متفاوت در معبر مورد مطالعه شناسایی شده که به کمک آن می‌توان تفسیر ساده‌تری از شلوغی و یا روان بودن آمد و شد در هر ساعت از روز داشت و قوانینی برای تعیین وضعیت ترافیک در آینده در قالب عبارت‌های کلامی ارائه کرد. مدل شبکه عصبی مصنوعی توسعه داده شده است که با دریافت مقدار متغیرهای سرعت و جریان در سه ساعت گذشته، وضعیت ترافیک را با بیش از ۷۵ درصد دقت برای داده‌های آزمایش تعیین می‌کند.

واژه‌های کلیدی: وضعیت ترافیک، داده‌کاوی، خوشه‌بندی، مهندسی ترافیک، راه‌های برون شهری

۱. مقدمه

در سال‌های اخیر موضوع جمع‌آوری و تحلیل اطلاعات پیرامون حجم ترافیک و تعیین وضعیت ترافیک در بزرگراه‌ها به مسئله‌ای جدی برای بسیاری از متولیان حمل‌ونقل تبدیل شده است. افزایش چشم‌گیر خودروهای شخصی، افزایش تقاضا برای سفرهای درون و برون شهری و بالا رفتن حجم ترافیک در بزرگراه‌ها و کلان‌شهرها، ساکنان کلان‌شهرها را با مشکلات عمده‌ای از قبیل افزایش زمان‌های سفر، کاهش سطح ایمنی، افزایش هزینه‌های مصرف سوخت و آلودگی‌های زیست محیطی روبرو ساخته است.

امروزه اطلاعات ترافیکی مربوط به آمد و شد انواع وسایل نقلیه مانند سرعت و جریان، یکی از اساسی‌ترین و مهم‌ترین داده‌های مورد نیاز برای مدیریت راه‌ها، برنامه‌ریزی برای کنترل و نظارت بر عبور و مرور وسایل نقلیه، تعیین زیرساخت‌ها و امکانات مورد نیاز در شبکه جاده‌ای، گذرگاه‌های درون شهری و برون‌شهری و تعمیر و نگهداری راه‌ها است. سازمان‌های متولی حمل‌ونقل در کشورها، در راستای توسعه مدیریت حمل‌ونقل در بزرگراه‌ها، با استفاده از دستگاه‌های ثبت اطلاعات خودکار عبور و مرور وسایل نقلیه که بخشی از تکنولوژی‌های سیستم‌های هوشمند حمل و نقل می‌باشند، اقدام به ثبت همزمان چندین مشخصه ترافیکی از جمله تعداد تردد در ساعت (جریان) و متوسط سرعت انواع وسایل نقلیه به تفکیک ساعات شبانه‌روز می‌کنند. در حال حاضر تعدادی از بزرگراه‌های شهر تهران من‌جمله معابر ورودی و خروجی اصلی شهر تهران به سامانه هوشمند تردد شمار مجهز شده است. این سامانه‌ها به طور مستمر، آمار حجم و سرعت عبور انواع وسایل نقلیه‌های سبک و سنگین ورودی و خروجی از تهران را به طور خودکار ثبت می‌کنند. اما جمع‌آوری داده‌ها به تنهایی دانش پنهان درون داده‌ها و به کارگیری از آنها جهت مدیریت ترافیک را فراهم نمی‌کند. لذا نیاز به توسعه روش‌ها و ابزارهایی جهت تسهیل این امر و تعیین وضعیت ترافیک و همچنین پیش‌بینی وضعیت آینده ترافیک در آینده نزدیک است. با تعیین وضعیت جاری و پیش‌بینی آینده وضعیت ترافیک بر اساس مشخصه‌های لحظه‌ای ثبت‌شده در سامانه‌های ثبت این اطلاعات، برنامه‌ریزان حمل‌ونقل قادر خواهند بود علاوه بر انجام تمهیدات لازم جهت تسهیل در عبور و مرور مانند حضور فیزیکی پلیس در این معابر و فراهم‌کردن وسایل امدادی، وضعیت ترافیک در یک مسیر را به اطلاع رانندگان برسانند تا رانندگان با مدیریت سفرهای خود به کاهش ترافیک در مسیر مورد نظر کمک کنند.

داده‌کاوی از جمله ابزاری است که تکنیک‌های مختلفی از جمله شبکه‌های عصبی[1]، قواعد انجمنی[2]، تجزیه و تحلیل رگرسیون، خوشه‌بندی، الگوریتم ژنتیک و ... را برای استخراج دانش درون داده‌ها به کار می‌گیرد. به عبارت دیگر روش‌های داده‌کاوی به دو دسته کلی تکنیک‌های پیشبینانه و توصیفی تقسیم می‌شوند که هدف تکنیک‌های توصیفی مانند خوشه‌بندی و قواعد انجمنی، توصیف یک رویداد یا واقعیت است در حالیکه تکنیک‌های پیشبینانه مانند رگرسیون، سری‌های زمانی و شبکه عصبی، به دنبال پیش‌بینی متغیرهای ناشناخته از داده‌های آتی می‌باشند. در این مقاله با استفاده از داده‌های ترافیکی ثبت شده در سامانه‌های مشخصه‌های ترافیکی مربوط به یکی از بزرگراه‌های ورودی به شهر تهران (بزرگراه ساوه-تهران) و به کمک تکنیک خوشه‌بندی به عنوان یکی از تکنیک‌های داده‌کاوی، نسبت به توصیف داده‌های ترافیکی و استخراج قواعد و روشی برای تعیین وضعیت ترافیک در یک مسیر اقدام می‌شود. همچنین از رویکرد شبکه‌های عصبی مصنوعی به منظور تعیین وضعیت ترافیک در ساعات بعدی بر اساس مقدار متغیرهای لحظه‌ای ترافیک در حال حاضر استفاده شده است. خروجی‌های این مدل این امکان را می‌دهد تا از طریق نمایشگرها و سیستم‌های اطلاعاتی، پیش‌بینی شرایط پرازدحام و حاد ترافیکی به رانندگان و برنامه‌ریزان حمل‌ونقل اطلاع‌رسانی شود تا نسبت به مدیریت ترافیک در مسیر مورد نظر اقدام شود.

۲. تعریف مسأله و اهداف تحقیق

وضعیت ترافیک در یک مسیر به وسیله متغیرهای ترافیکی مختلفی تعیین می‌شود. اما از آنجائیکه ترافیک مقوله‌ای پیچیده و متأثر از عوامل بسیار زیادی است و متغیرهای ترافیکی دائماً در حال تغییراند، هیچ متغیر ترافیکی به تنهایی نمی‌تواند تمام جوانب وضعیت آن را مشخص کند. به عبارتی متغیرهای ترافیکی جنبه‌های مختلف وضعیت ترافیک در یک مسیر را اندازه‌گیری می‌کنند و به تنهایی وضعیت ترافیک را نشان نمی‌دهند. برای بیان وضعیت ترافیک و شلوغی در یک مسیر نیاز است تا حداقل دو متغیر، به طور همزمان اندازه‌گیری و تحلیل شوند. برای مثال اگر تعداد وسیله نقلیه عبوری بر واحد ساعت، متغیر جریان، مقداری کم و یا زیاد باشد نمی‌توان در مورد شلوغی و یا خلوتی مسیر اظهار نظر کرد. چراکه در هر دو حالت بسته‌بودن و خلوتی مسیر این متغیر، مقدار کمی خواهد داشت. اما چنانچه متغیر چگالی یا سرعت وسایل نقلیه نیز همزمان اندازه‌گیری شوند، با تحلیل همزمان هر دو متغیر مانند جریان و سرعت، امکان تعیین وضعیت شلوغی، ازدحام و روانی ترافیک در مسیر فراهم خواهد شد. از طرف دیگر بیان کیفی وضعیت ترافیک در یک مسیر و یا یک متغیر ترافیکی در قالب عبارت‌های «کم» یا «زیاد» بر مبنای مقدار عددی آن، بستگی به مشخصات مسیر مانند روسازی، روشنایی و ... دارد و نمی‌توان یک بازه برای تمامی مسیرها ارائه داد. بنابراین تغییر دائم متغیرهای ترافیکی، وابستگی آنها به یکدیگر و ناتوانی در تفسیر وضعیت ترافیک به علت اختلاط نامناسب داده‌ها، این تحقیق را به سمت استفاده از تکنیک‌های داده‌کاوی پیش برده است. هر چند تحقیقات محدودی در گذشته مانند هی و همکاران برای پیش‌بینی مقدار متغیرهای ترافیکی مثل جریان با استفاده از داده‌کاوی انجام شده است، اما فعالیت آنها در راستای پیش‌بینی مقدار متغیرهای ترافیکی بوده و پاسخی به مسئله تعیین وضعیت ترافیک نبوده است. علاوه بر این، سیستم‌های حمل و نقل هوشمند[3] (ITS)، حجم زیادی از اطلاعات ترافیکی را در ساعت جمع‌آوری و ذخیره می‌کنند. بنابراین نیاز ضروری به روش‌هایی برای کاوش دانش پنهان در این پایگاه داده‌ها احساس می‌شود. به نقل از زمانی و همکاران داده‌های ذخیره شده در پایگاه داده‌های جهان هر بیست ماه ۱۰۰٪ افزایش می‌یابند و با وجود رشد بالای داده‌ها، استخراج دانش از آنها بسیار ضعیف می‌باشد. بنابراین داده‌کاوی به عنوان یک ابزار مفید برای دستیابی به اطلاعات، به منظور تنظیم استراتژی‌های مناسب و یافتن فرصت‌های جدید، نقش موثری خواهد داشت.

خوشه‌بندی با استفاده از تجزیه و تحلیل خوشه‌ای که یکی از برجسته‌ترین روش‌های داده‌کاوی در سال‌های اخیر بوده و در بسیاری از نرم‌افزارهای آماری قابل استفاده می‌باشد، مشاهدات را به خوشه‌های همگنی تقسیم می‌کند به گونه‌ای که داده‌های درون هر خوشه بیشترین مشابهت را از نظر متغیرهای لحاظ شده در خوشه‌بندی داشته باشند و بیشترین تمایز بین خوشه‌ها اتفاق افتد. در این مقاله، از میان ابزارهای موجود، الگوریتم K-Means با هدف تعریف خوشه‌هایی با الگوهای ترافیکی یکسان به منظور درک ویژگی‌های منحصر به هر خوشه و ایجاد آمادگی برای تهیه زیرساخت‌های لازم در شرایط بحرانی، بر روی داده‌های مربوط به بزرگراه ورودی ساوه–تهران در بازه دوساله به کار گرفته شده است. بنابراین هدف این مقاله ارائه تکنیکی برای تفسیر وضعیت ترافیکی در بزرگراه‌های درون و برون شهری بر اساس داده‌های ترافیکی یک مسیر و با استفاده از رویکرد خوشه‌بندی می‌باشد. بر اساس این تکنیک، ساعات مورد مطالعه بر اساس دو متغیر جریان و متوسط سرعت در یک ساعت خوشه‌بندی شده و به این ترتیب، ساعاتی از روز که دارای متغیرهای ترافیکی مشابه در مسیر مورد مطالعه می‌باشند در یک خوشه قرار گرفته‌اند. وضعیت ترافیکی مسیر در هر ساعت، بر اساس تغییرات متغیرها در هر خوشه تحلیل می‌شوند. بر اساس رویکرد پیشنهادی، مدیران حوزه ترافیک قادر خواهند بود قواعدی برای کشف وضعیت ترافیک در مسیر موردنظر در اختیار داشته باشد. همچنین پیش‌بینی وضعیت ترافیک بر اساس قواعد استخراج شده از الگوریتم خوشه‌بندی بر اساس عبارت‌های کلامی مناسب که گویای وضعیت ترافیک در یک مسیر باشد به

مدیریت ترافیک در مسیر کمک خواهد کرد. در این مقاله با استفاده از رویکرد شبکه‌های عصبی پیش‌خوراند مدلی طراحی شده که مقدار متغیرهای ترافیکی مسیر در یک ساعت را به عنوان ورودی دریافت می‌کند و وضعیت ترافیک (شماره خوشه) را در ساعت بعدی پیش‌بینی می‌کند.

۳. روش تحقیق

روش تحقیق در این مقاله بر اساس متدولوژی CRISP-DM[۴] (یکی از متدولوژی‌های داده‌کاوی) اجرا شده است. طبق این مدل گام های اصلی در فرایند داده‌کاوی به شرح زیر می‌باشد.

۳-۱ درک فضای کسب و کار(شناخت سیستم ترافیک)

مدیریت ترافیک براساس داده‌های ترافیکی، رویکرد مورد توجه بسیاری از تحقیقات می‌باشد. برای تشخیص دقیق وضعیت و حجم ترافیک در ساعات مختلف روزهای عادی، تعطیلات و روزهای پایانی هفته، تردد شمارهایی در مبادی ورودی و خروجی شهر تهران نصب شده است تا از این طریق میزان ورود و خروج خودروها به شهر مشخص شود و این کار از جمله فعالیت‌هایی است که در تلاش برای بررسی ترافیک پایتخت به کار گرفته شده و از نیمه دوم سال ۸۷ شروع به کار کرده است. آمار تردد موجود در هر محور ابزار مطلوبی برای برنامه‌ریزی و مدیریت ترافیک در مسیرها به تفکیک شرایط است. ترددشمار آمارهایی درباره نحوه توزیع ترافیک در شبکه، تعیین اهمیت نسبی محورهای شبکه، اولویت‌بندی طرح‌های ساخت و توسعه و ارزیابی اقتصادی و نظارت بر عملکرد سیستم حمل‌ونقل به دست می‌دهد که ارائه‌کننده اطلاعات کاربردی و ارزشمندی بالایی است. علاوه بر این در دسترس بودن داده‌های مربوط به سرعت خودروهای عبوری به تعریف دقیق‌تری از وضعیت ترافیک در هر ساعت از روز منجر خواهد شد.

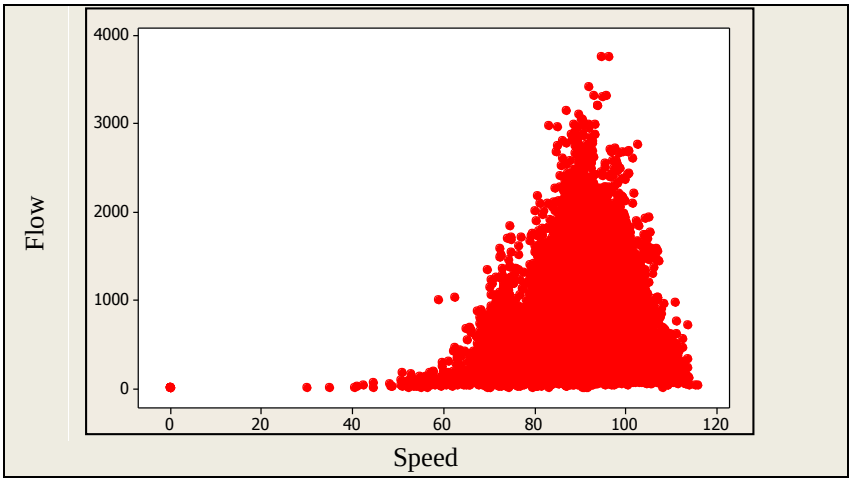
۳-۲ انتخاب، شناخت و آماده‌سازی داده‌ها

در این بخش داده‌های مربوط به تحقیق جمع‌آوری، پاکسازی و آماده‌سازی شده‌اند. در سامانه ثبت داده‌های تردد، به ازای هر خودروی عبوری از حسگرهای ثبت داده‌ها، لحظه عبور شامل تاریخ، ساعت، دقیقه و ثانیه عبور در یک فیلد، سرعت وسیله نقلیه، نوع وسیله نقلیه (موتور سیکلت، سواری، اتوبوس و انواع کامیون) به همراه خط عبوری از مسیر در فیلدهای جداگانه‌ای ثبت می‌شود. به منظور محاسبه متغیرهای جریان (تعداد خودروی عبوری از معبر در واحد ساعت) و متوسط سرعت (متوسط سرعت لحظه‌ای خودروهای عبوری در یک ساعت) برای ورودی ساوه – تهران در بازه دو سال، از نرم افزار SQL استفاده شده است. تفکیک داده‌های مربوط به تاریخ از فیلد مربوطه و طراحی و اجرای پرس و جوهای لازم جهت محاسبه دو متغیر ترافیکی جریان و متوسط سرعت در یک ساعت از فعالیت‌های مربوط به آماده‌سازی داده‌ها در این مرحله می‌باشد. در بازه جمع‌آوری داده‌ها بیش از هفده هزار مشاهده (ساعت مورد مطالعه) در پایگاه داده‌ای که مربوط به عبور بیش از ۱۵ میلیون خودرو می‌باشد، ثبت شده است.

۳-۳ مدلسازی (خوشه‌بندی)

روش‌های آنالیز خوشه‌بندی با داشتن توان بالایی برای شناسایی رفتارهای مشابه در مجموعه داده‌های پیچیده به آشکارکردن الگوهای پنهان در این داده‌ها کمک خواهد نمود. در میان الگوریتم‌های خوشه‌بندی، الگوریتم-K میانگین به دلیل سرعت، دقت و سادگی بالا مورد توجه قرار گرفته است. این الگوریتم تلاش می‌کند با خوشه‌بندی داده‌ها در K خوشه (K مقداری ثابت است) مقدار مربع فاصله هر یک از مشاهدات را از خوشه تخصیص یافته به آن کاهش دهد. تعیین مراکز دسته در هر خوشه به صورت تصادفی است و این روند تا زمان عدم تغییر در مراکز خوشه‌ها، به صورت متوالی ادامه می‌یابد و در نهایت K خوشه با بیشترین مشابهت در داده‌های درون هر خوشه و بیشترین تفاوت بین خوشه‌ها حاصل می‌شود.

گام اول برای استفاده از الگوریتم K-Means، انتخاب مشخصه‌ها یا متغیرهایی است که قرار است در الگوریتم خوشه‌بندی به کار گرفته شوند. برای این انتخاب سعی شده است از مشخصه‌هایی استفاده شود که بتوان ویژگی‌های داخل هر خوشه و وضعیت ترافیک را با کمک آن تحلیل کرد. معروف‌ترین متغیرهای توصیف جریان ترافیک، تعداد وسیله نقلیه‌ای که از یک نقطه ثابت عبور می‌کنند و متوسط سرعت وسایل نقلیه در یک بازه زمانی معین مانند یک ساعت می‌باشد. بنابراین در این مقاله، این دو متغیر برای تحلیل وضعیت ترافیکی خوشه‌ها انتخاب شده‌اند. نمودار پراکندگی دو متغیر جریان و سرعت نسبت به یکدیگر به ازای داده‌های جمع‌آوری شده، در شکل زیر آمده است. همانطور که مشخص است دامنه تغییرات متوسط سرعت بین ۰ تا ۱۲۰ کیلومتر در ساعت و جریان بین صفر تا حدود ٤٠٠٠ خودرو در ساعت متغیر است. مشاهده مبدا مختصات مربوط به ساعاتی هستند که ترددی ثبت نشده یا خودروها به دلایلی متوقف شده‌اند. تشخیص وضعیت ترافیک با مطالعه این نمودار به دلیل پراکندگی زیاد دو متغیر امکان‌پذیر نمی‌باشد و نیاز به ابزارهایی مانند رویکرد پیشنهادی این تحقیق بیش از پیش احساس می‌شود.



شکل ۱. نمودار پراکندگی دو متغیر جریان و سرعت نسبت به یکدیگر

برای تعیین وضعیت ترافیک در یک مسیر بر اساس دو متغیر نرخ جریان (Flow) و سرعت (Speed) می‌توان به بررسی حالات مختلف بین این دو متغیر پرداخت.

حالت اول: ساعتی از روز که هم نرخ جریان و هم سرعت، هر دو مقادیری بیشتر از حد متعارف داشته باشند، البته حدود نرمال دو متغیر بر اساس داده‌های گذشته در هر مسیر تعیین می‌شود و تابعی از مشخصه‌های مسیر است. در این حالت با وجود مسیر شلوغ و پرتردد، جریان ترافیکی روان بوده و خودروهای مسیر با حداکثر سرعت در حال حرکت‌اند.

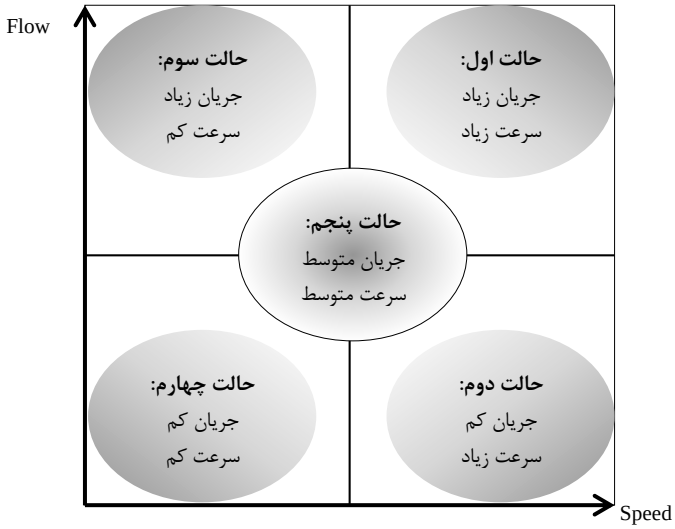
حالت دوم: ساعتی از روز که هر دو متغیر نرخ جریان و سرعت مقادیر پائینی دارند. در این حالت مسیر تقریباً مسدود می‌باشد و این وضعیت، ایجاد ترافیک و شرایط بحرانی نموده است. در این حالت، دو وضعیت خاص را نیز می‌توان متصور شد. ابتدا اینکه به علت خاصی مسیر کاملا مسدود شده است به عبارتی در یک ساعت، در مسیر خودروها به طور کامل متوقف شده‌اند. اتفاق دوم زمانی است که در یک ساعت هیچ خودرویی از مسیر عبور نکرده و مسیر کاملا خالی از هر خودرویی است. در این دو وضعیت، متوسط سرعت و جریان به مقدار صفر می‌رسد. از آنجاییکه این دو وضعیت، از حالات خاص به شمار می‌رود و احتمال وقوع آنها به خصوص در بزرگراه مورد مطالعه بسیار پایین است، تفکیک این دو وضعیت از یکدیگر به مطالعه بیشتر متغیرهای ترافیکی در مسیر مورد نظر دارد. به عنوان مثال اگر در ساعت بامداد چنین حالتی رخ دهد، احتمال اینکه ترافیک در وضعیت خالی بودن مسیر باشد زیادتر است و یا اینکه اگر در

ساعات میانی روز متغیرهای ترافیکی در این حالت قرار گیرند و تردد در ساعات قبل از آن زیاد باشد احتمال قرار گرفتن در وضعیت توقف وسایل در یک ساعت به دلیل تصادفات زیاد است.

حالت سوم: مربوط به ساعتی است که نرخ جریان زیاد اما سرعت پایین است. در این حالت شلوغی و پرتردد بودن مسیر موجب کاهش سرعت متوسط خودروها شده است. اما از آنجاییکه نرخ جریان مقدار بالایی دارد می‌توان دریافت که ترافیک سنگینی به وجود نیامده و جریان ترافیک روان است.

حالت چهارم: جریان کم، اما سرعت زیاد و در نتیجه این حالت نشان دهنده مسیر آزاد و خلوت است.

حالت پنجم: حالت میانه‌ای نیز وجود دارد که سرعت و جریان در یک ساعت پیرامون حدود متوسط متغیر تغییر می‌کنند که در این شرایط وضعیت ترافیک عادی است. البته وضعیت عادی به معنای وضعیت ایده‌ال ترافیک در یک مسیر نیست بلکه به معنای وضعیتی است که در غالب اوقات، مسیر مورد نظر همان وضعیت را داراست و بیشترین احتمال وقوع را دارد. حالات ذکر شده، در شکل ۲ قابل مشاهده است.



شکل ۲. حالات مختلف بین دو متغیر جریان و سرعت

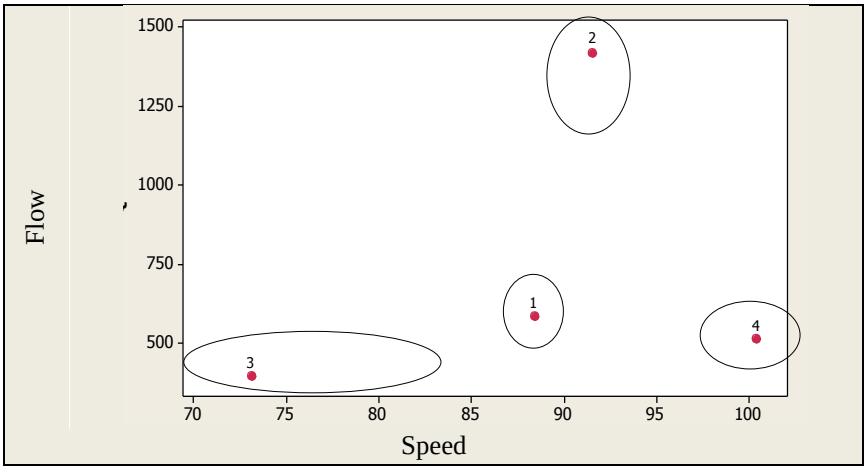
پس از انتخاب متغیرهای بخش‌بندی (جریان و سرعت در یک ساعت)، الگوریتم K-Means در نرم‌افزار Matlab به ازای تعداد خوشه‌های مختلف اجرا شده و شاخص دیویس بولدین[۵] (یکی از شاخص‌های ارزیابی کیفیت خوشه‌بندی است که هر چه مقدار آن کمتر باشد بیانگر کیفیت بالاتر خوشه‌بندی می‌باشد) برای یافتن تعداد خوشه بهینه مورد استفاده قرار گرفت که نتایج آن در جدول ۱ قابل مشاهده است. همانگونه که مشاهده می‌شود براساس کمترین میزان شاخص دیویس بولدین، تعداد خوشه انتخابی، ٤ خوشه انتخاب شده است.

جدول ۱. مقادیر شاخص دیویس بولدین به ازای تعداد خوشه‌های مختلف

تعداد خوشه	۲	۳	٤	٥	٦
شاخص دیویس بولین	۱/۲۸٤	۰/۹٦۰	۰/۷٦٤	۰/۹۱۱	۰/۸۳۳

حال با توجه به اینکه تعداد خوشه مناسب برای اجرای الگوریتم K-Means انتخاب شده است می‌توان با انجام فرایند خوشه‌بندی، شماره خوشه هر یک از مشاهدات (ساعت مورد مطالعه) را تعیین کرد. بخش بعد با بررسی وضعیت هر یک از خوشه‌ها در دو متغیر نرخ جریان و سرعت، اطلاعات بیشتری راجع به خوشه‌ها، فراوانی مشاهدات در آنها، ویژگی‌های منحصر

به فرد آنها، عوامل تأثیرگذار بر هر خوشه و ... استخراج خواهد شد. پس از انجام خوشه‌بندی، میانگین سرعت و جریان ساعات هر خوشه در شکل ۳ نمایش داده شده‌اند. در بخش قبلی برای حالات بین سرعت و جریان، پنج وضعیت متصور شده بود. در ادامه باید مشاهده کرد که خوشه‌ها در کدام یک از وضعیت‌های بالا قرار خواهند گرفت. همانگونه که در این شکل مشاهده می‌شود خوشه ۱ بیانگر حالت پنجم (سرعت و جریان متوسط)، خوشه ۲ نزدیک به حالت اول (جریان زیاد و سرعت بالا)، خوشه ۳ بیانگر حالت چهارم (جریان کم و سرعت پایین) و خوشه ٤ نزدیک به حالت دوم (جریان کم و سرعت بالا) می‌باشد.



شکل ۳. میانگین سرعت و نرخ جریان در هریک از خوشه‌های انتخابی

حال موضوع مهم این است که دسته‌بندی مشاهدات در این خوشه‌ها با چه قوانینی صورت پذیرفته است. به عبارتی چه تعداد از مشاهدات در هر یک از خوشه‌ها قرار گرفته و این مشاهدات چه رفتار یا ویژگی مشابهی با هم داشته‌اند؟ آیا این دسته‌بندی ما را به سمت استخراج قوانینی از رفتار مشاهدات در این بزرگراه هدایت خواهد کرد؟ در ادامه به کمک تحلیل‌های توصیفی و ترافیکی به این سوالات پاسخ داده خواهد شد.

٤-۳ تحلیل توصیفی و ترافیکی خوشه‌ها

جدول ۱ خلاصه آماری هر یک از متغیرهای ترافیکی را به تفکیک هر خوشه نشان می‌دهد. با بررسی این جدول می‌توان موارد زیر را برداشت کرد:

خوشه۱: ترددهایی با نرخ جریان پایین و سرعت متوسط رو به بالا را شامل می‌شود. این خوشه احتمالاً ترکیبی از مشاهدات پیک و مشاهدات عادی می‌باشد اما از آنجاییکه ۳٦ درصد از مشاهدات را شامل می‌شود احتمالاً این خوشه نشان‌دهنده مشاهدات ساعات عادی روز است و اگر بخواهیم این خوشه را در دسته‌بندی تعریف شده در شکل ۲ جای دهیم، این خوشه در حالت پنجم (حالت متوسط) قرار خواهد گرفت.

خوشه۲: مشاهده شده با وجود نرخ جریان بالا در مشاهدات، سرعت عبور و مرورها نیز بالاتر از حد متوسط می‌باشد بنابراین می‌توان نتیجه گرفت مشاهدات موجود در این خوشه، مربوط به ساعاتی است که حجم عبور و مرور زیاد بوده اما منجر به ترافیک نشده و مسیر باز است و یا به عبارتی حجم زیادی از خودروها در حال عبور از معبر با حداکثر سرعت می‌باشند.

خوشه۳: با کمترین نرخ جریان در ساعت و پایین‌ترین سرعت متوسط، خوشه‌ای با پیک ترافیک است. علاوه بر این، تعداد مشاهدات این خوشه تنها ۸ ٪ از کل مشاهدات است.

خوشه٤: نیز که در آن متوسط سرعت بالا و نرخ جریان پایین است بیانگر حالات عادی و عاری از هر گونه ترافیک می‌باشد. همانگونه که مشاهده می‌شود حدود ۱٦ ٪ از مشاهدات در این خوشه جای گرفته‌اند.

جدول ۲. خلاصه شاخص‌های آماری خوشه‌ها

درصد فراوانی خوشه	واریانس جریان	میانگین جریان	واریانس سرعت	میانگین سرعت	وضعیت ترافیک در شکل ۲	شماره خوشه
٪۳٦	۲٦٦/٤۸۰	۵۸۳/٤۰٤	۳/٤۲۲۰۳	۸۸/٤۲۰۳	حالت پنجم	۱
٪٤۰	۳۱٦/۵٤۹	۱٤۱٦/۷۷	۳/۸۷٦۷۵	۹۱/۵۷۹۳	حالت اول	۲
٪۸	۲۸۸/۳۰۷	۳۹۲/۷٦۹	۸/٦۰٤۲۵	۷۳/۱۱۵۱	حالت چهارم	۳
٪۱٦	۲۸۸/۵۵۲	۵۱۲/۰٦۸	٤/٤۲۵٦۱	۱۰۰/٤۳۳	حالت دوم	٤

مشاهدات استفاده شده در خوشه‌بندی شامل مشاهداتی از جنس تمام انواع خودروها می‌باشد. بنابراین با مشاهده اختلاف موجود در مقدار متوسط سرعت و نرخ جریان در وسایل نقلیه سواری و سایر انواع وسیله نقلیه و با توجه به بالا بودن تعداد وسایل نقلیه سواری نسبت به سایر انواع، ممکن است تصور شود که نتایج خوشه‌بندی بر اساس خودروهای سواری، با نتایج حاصل از خوشه‌بندی با ادغام کل داده‌ها متفاوت خواهد بود. بنابراین یک بار دیگر با استفاده از همان روش خوشه‌بندی، الگوریتم اجرا شده و خوشه‌های زیر با متوسط سرعت و نرخ جریان جدید حاصل شد. واضح است که در این حالت نیز نتایج تفاوتی با حالت قبل ندارند بنابراین مبنای تحلیل ترافیکی، همان خوشه‌بندی سابق می‌باشد.

جدول ۲. متوسط نرخ جریان و سرعت متوسط در خوشه‌بندی بر مبنای خودروهای سواری

میانگین سرعت	میانگین جریان	شماره خوشه
۱۰۰/٤	٤۰۰/٤	۱
۹٤/۱۷	۱۱۷٤/۲	۲
۷۱/۷۵	۲۸٤/۹	۳
۸۹/۳۵	٤٦۳/۷	٤

جدول زیر میانگین جریان و میانگین سرعت را در انواع وسایل نقلیه به ازای هر خوشه نشان می‌دهد. درصد وجود وسایل نقلیه سنگین در هر خوشه می‌تواند به توجیه متوسط سرعت و متوسط میانگین جریان کمک کند. همانطور که در جدول ٤ مشاهده می‌شود بیشینهٔ نرخ جریان در تمام انواع وسیله نقلیه اعم از سواری، تریلر، اتوبوس و کامیون، در خوشه ۲ رخ داده است. بعد از بالاترین متوسط نرخ جریان در خوشه ۲، دومین رتبه مربوط به خوشه‌های ۱ و ٤ می‌باشد (در خودروهای سواری(۱۱) که نرخ جریانی شبیه به یکدیگر دارند. بررسی‌ها نشان می‌دهد متوسط سرعت در خوشه ۱ نزدیک به ۱۰ کیلومتر بر ساعت از متوسط سرعت در خوشه ٤ کمتر است (۱۰۲ در مقابل ۹۱) و علت آن را می‌توان درصد وسایل نقلیه سنگینی دانست که در خوشه ۱ به طور چشمگیری از خوشه ٤ بیشترند و این موضوع، قطعاً بر سرعت متوسط خودروهای سواری تاثیرگذار بوده است.

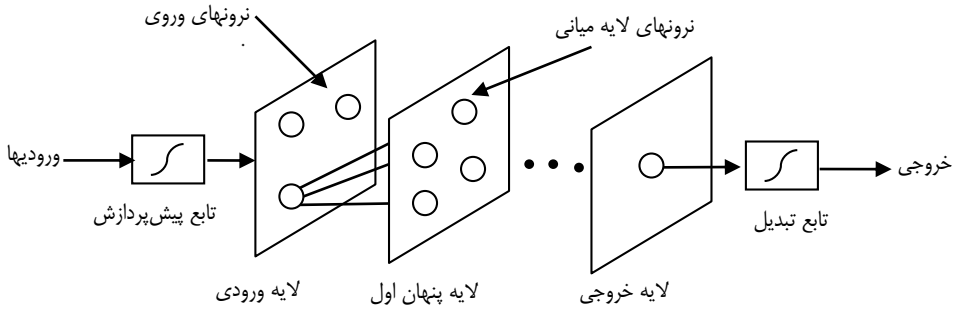
جدول ۳. نرخ جریان و سرعت متوسط در وسایل نقلیه وخوشه‌ها

نوع خودرو	سواری		تریلر		اتوبوس		کامیون	
شماره خوشه	Flow	Speed	Flow	Speed	Flow	Speed	Flow	Speed
۱	٤۳۲/۲	۹۱	۷۵/٤	۸۰/٤	۲٦/۸	۸۱/۷	۱٤/۷	۸۱/۲
۲	۱۱٦٤/٦	۹۳/۹	۱۲۷/٦	۸۰/۵	٤۳/۸	۸۱	۲۱/۲	۸۰/٦
۳	۲۹۱	۷۲/۸	٤۸/۵	۷۰/۷	۲۳	۷۲/٦	۱۲/۳	۷۳/۸
٤	٤٤۵/۳	۱۰۲/۵	۳۲/٦	۸٦/۳	۱٤	۸٦/۲	٦/۲	۸۷/۳

با بررسی متوسط سرعت در خودروهای سنگین، مشاهده شد این مقدار در تمام خوشه‌ها نزدیک به ۸۰ کیلومتر بر ساعت می باشد مگر در خوشه ۳ که به مقدار ۷۰ کیلومتر بر ساعت رسیده است. در مورد خودروهای سواری نیز، متوسط سرعت و البته متوسط نرخ جریان، کمترین مقدار را در بین خوشه‌ها در خوشه ۳ داشته است. بنابراین می‌توان نتیجه گرفت در خوشه ۳، کاهش سرعت هم در خودروهای سواری و هم سنگین، به علت بسته بودن مسیر ایجاد شده است. از آنجائیکه تعداد مشاهدات این خوشه تنها ۸٪ از کل داده‌ها می‌باشد، احتمال وجود تصادفات و یا سایر عواملی که منجر به بسته شدن مسیر می‌شود، در این خوشه زیاد است. برای یافتن داده‌های اطلاعاتی و استخراج قوانینی از مشاهدات هر خوشه نیاز است نسبت به عوامل تاثیرگذار بر هر خوشه اطلاعاتی کسب کرد.

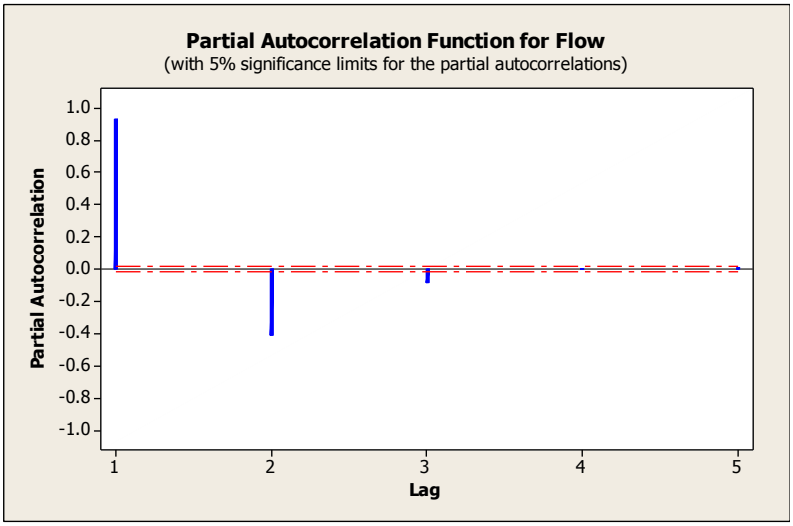
۳-۵ پیش‌بینی وضعیت ترافیک (توسعه نتایج)

تعیین وضعیت حال حاضر ترافیک و یا ساعات گذشته راه‌کاری برای حل مشکلات ترافیک در یک مسیر نمی‌باشد. بلکه پیش‌بینی وضعیت ترافیک در ساعات آینده به مدیریت ترافیک کمک خواهد کرد. در این بخش مدل پیش‌بینی مبتنی بر شبکه‌های عصبی مصنوعی توسعه یافته است که با دریافت متغیرهای ورودی مناسب، وضعیت ترافیک (شماره خوشه) را در ساعت آینده تعیین می‌کند. شبکه‌های عصبی مصنوعی با پردازش روی داده‌های تجربی، دانش یا قانون نهفته در ورای داده‌ها را به ساختار شبکه منتقل می‌کند که به این عمل یادگیری می‌گویند. در این مدل نرون‌ها کوچکترین واحد پردازش اطلاعات می‌باشند. هر شبکه از یک لایه ورودی، یک لایه خروجی و تعدادی لایه میانی با تعدادی نرون درون هر لایه تشکیل شده است و نرون‌های هر لایه با وزنهایی به نرونهای لایه بعد متصل شده‌اند. طی فرآیند آموزش این وزن‌ها جهت دستیابی به شبکه‌ای با خطای پیش‌بینی کمتر پیوسته تغییر می‌کنند. همچنین خروجی‌های هر نرون بعد از عبور از توابع تبدیل به نرون‌های لایه بعد منتقل می‌شوند. شکل ٤ اجزای یک شبکه عصبی پیش خوراند را نشان می‌دهد.



شکل٤. اجزاء یک شبکه عصبی مصنوعی پیش‌خوراند

ورودی‌های شبکه عصبی مصنوعی برای پیش بینی وضعیت ترافیک ساعت آتی می‌تواند متغیرهای ترافیکی در حال حاضر یا ساعات گذشته باشد. جهت تعیین تعداد دوره‌های وابستگی وضعیت ترافیک به متغیرهای ترافیک در گذشته، نمودار خود همبستگی جزئی[٦] (PAF) برای متغیر جریان به صورت شکل ۵ رسم شده است.



شکل۵. نمودار خودهمبستگی جزئی متغیر جریان

نمودار خود همبستگی جزئی (شکل۵) نشان می‌دهد، مقدار جریان حداکثر به مقدار جریان در سه دوره (ساعت) گذشته وابسته است. لذا برای پیش‌بینی وضعیت ترافیک (شماره خوشه) در ساعت آینده نیاز به دانستن متغیرهای ترافیکی سرعت و جریان ساعت جاری و دو دوره قبل می‌باشد. بر این اساس ورودی‌های شبکه عصبی طراحی شده شش متغیر می‌باشد (جریان و سرعت در سه ساعت قبل از دوره پیش‌بینی) و خروجی این شبکه شماره خوشه یا وضعیت ساعت آینده ترافیک می‌باشد.

جهت ساخت و تست عملکرد مدل شبکه عصبی، داده‌های جمع‌آوری شده به همراه نتایج خوشه‌بندی در پایگاه داده‌ای قرار داده شده و جهت حفظ سری زمانی مناسب، داده‌ها پاکسازی و مرتب شده‌اند. داده‌ها شامل ساعات مختلف روزهای مورد مطالعه مشتمل بر ۱۷۵۲۹ ساعت می‌باشد که به دو گروه داده‌های آموزش و آزمایش به ترتیب به نسبت ۸۰ (۱٤۰۲۱مشاهده) به ۲۰ (۳۵۰۵ مشاهده) به صورت تصادفی تقسیم شده‌اند. از داده‌های آموزش برای ساخت مدل شبکه عصبی و از داده‌های آزمایش برای ارزیابی دقت پیش‌بینی استفاده شده است. شبکه

طراحی شده علاوه بر یک لایه خروجی و شش لایه ورودی، دارای دو لایه میانی با ۲۰ نرون درون هر لایه می‌باشد. پس از آموزش شبکه، مشاهدات آزمایش (متغیرهای ورودی به ازای هر ساعت داده) به شبکه وارد شده و شماره خوشه توسط شبکه پیش‌بینی شده است. درصد انطباق شماره خوشه پیشنهادی توسط شبکه عصبی با شماره واقعی خوشه که توسط مدل خوشه‌بندی حاصل شده است، به عنوان معیار عملکردی مدل درنظر گرفته شده است. در جدول ۵، درصد انطباق شماره خوشه پیش‌بینی شده و روش خوشه‌بندی K-Means نشان داده شده است.

جدول ۵. ارزیابی نتایج خوشه‌بندی با نمودار شبکه عصبی

نوع داده‌ها	داده‌های آموزش		داده‌های آزمایش	
	درصد	تعداد	درصد	تعداد
تعداد تشخیص خوشه صحیح	۷۶/۳	۱۰۶۹۱	۷۵/۷	۲۶۵٤
تعداد تشخیص خوشه اشتباه	۲۳/۷	۳۳۳۰	۲٤/۳	۸۵۰
مجموع	۱۰۰	۱٤۰۲۱	۱۰۰	۳۵۰٤

نتایج جدول ۵ نشان می‌دهد که شبکه عصبی طراحی شده، ۷۶/۳ درصد مواقع شماره خوشه به کار رفته برای آموزش را به درستی پیش‌بینی می‌کند. همچنین بیش از ۷۵ درصد مواقع، مدل شبکه عصبی شماره خوشه داده‌های آزمایش (داده‌هایی که مدل شماره خوشه آنها را نمی‌داند) به درستی پیش‌بینی می‌کند. جدول ٦، درصد خطای مدل پیش‌بینی را به تفکیک خوشه‌ها نشان می‌دهد. به عنوان مثال ۷۱/۱ درصد مواقع مدل، پیش‌بینی شماره خوشه وضعیت عادی را درست پیش‌بینی نموده است و تنها ٤/۵ درصد شماره خوشه را به جای عادی، خوشه بحرانی (سرعت و جریان پایین، خوشه ۳) پیش‌بینی نموده است. همانطور که در جدول ٦ مشاهده می‌شود، شماره خوشه پیشنهادی توسط مدل خیلی با خوشه واقعی مشاهده تفاوت ندارد. بیشترین انحراف خوشه پیش‌بینی و خوشه واقعی در خوشه واقعی ٤ رخ داده است. لذا استفاده از مدل توسعه داده شده در این پژوهش با اندکی خطا می‌تواند به عنوان ابزاری برای پیش‌بینی وضعیت ترافیک در آینده باشد و با استفاده از سیستم‌های اطلاع‌رسانی مناسب وضعیت آینده ترافیک در قالب عبارت‌های کلامی مناسب که بیانگر چگونگی ازدحام خودروها در مسیر و روان بودن ترافیک باشد، در اختیار رانندگان و برنامه‌ریزان شهری و برون‌شهری قرار گیرد تا با انجام تدابیر مناسب اقدامی در راستای مدیریت ترافیک برداشته شود.

جدول ٦. درصد خطاهای پیش‌بینی خوشه به تفکیک خوشه‌ها

پیش‌بینی / واقعی	وضعیت عادی (خوشه ۱)	سرعت و جریان زیاد (خوشه ۲)	سرعت و جریان کم (خوشه ۳)	سرعت زیاد و جریان کم (خوشه ٤)
وضعیت عادی (خوشه ۱)	۷۱/۱	۲٤	٤/۵	۰/٤
سرعت و جریان زیاد (خوشه ۲)	۵/۵	۹۰/۱	٤/۳	۰/۱
سرعت و جریان کم (خوشه ۳)	۱۳	۱٦	۷۱	۰
سرعت زیاد و جریان کم (خوشه ٤)	۵/۲	۱٦/۳	۲۷	۵۱

٤. نتیجه‌گیری

تحلیل وضعیت ترافیک در یک مسیر با دانستن مقادیر متغیرهای ترافیکی مانند سرعت و جریان در آن لحظه آسان نیست بلکه

تعیین وضعیت ترافیک از جمله شلوغی مسیر و یا روان بودن جریان آن، بستگی به دانستن مقدار تغییرات متغیرهای ترافیکی در آن مسیر دارد. استفاده از رویکرد خوشه‌بندی امکان بررسی همزمان تغییرات دو متغیر ترافیکی (سرعت و جریان) و تعیین وضعیت ترافیک در مسیر را می‌دهد. در این مقاله با استفاده از داده‌های تاریخی معبر ساوه–تهران نشان داده شده که وضعیت ترافیک در ٤ خوشه با رفتارهای مشابه گروه‌بندی می‌شوند. تحلیل مقدار متغیرهای ترافیکی سرعت و جریان درون هر خوشه وضعیت ترافیک را از دو جنبه ازدحام و جریان ترافیک نشان می‌دهد. یکی از خوشه‌های شناسایی شده بیانگر وضعیت عادی ترافیک بیانگر است که در آن متغیرهای سرعت و جریان در حدود میانه خود در روزهای مختلف تغییر می‌کند. خوشه‌ای که متغیر جریان نسبت به میانگین زیاد ولی سرعت پایین است، ساعاتی را با وضعیت شلوغی مسیر اما روانی ترافیک در مسیر نشان می‌دهد. به همین ترتیب خوشه‌ای با وضعیت بحرانی ترافیک شناسایی شده که متغیرهای جریان و سرعت در آن پایین است. تحلیل سری زمانی داده‌های متغیر جریان در واحد ساعت نشان می‌دهد، متغیر جریان در هر ساعت تابع مقدار جریان در سه دوره (ساعت) گذشته آن می‌باشد. لذا مدل شبکه عصبی توسعه داده شده که با دریافت مقدار متغیرهای ترافیکی سرعت و جریان در سه دوره (ساعت) قبل شماره خوشه که بیانگر وضعیت ترافیک می‌باشد را پیش‌بینی می‌کند. اجرای مدل شبکه عصبی توسعه داده شده برای داده‌های آزمایش بیش از ۷۵ درصد دقت پیش‌بینی دارد. توسعه مدل‌های دیگر خوشه‌بندی به منظور افزایش دقت خوشه‌بندی و همچنین مدل‌های دقیق‌تر پیش‌بینی وضعیت ترافیک در آینده تحقیقات آتی این پژوهش می‌باشد.

۵. پی‌نوشت‌ها

1. Artificial Neural Network
2. Association Rules
3. Intelligent Transportation Systems
4. Cross-Industry Standard Process for Data Mining
5. Davies Bouldin
6. Partial Autocorrelation Function

٦. مراجع

– حاجی حسینلو، منصور و قائمی سید علی (۱۳۹۳) "ارزیابی تاثیر وسایل نقلیه سنگین بر جریان ترافیک و میزان انتشار آلاینده‌های ناشی از آن در شبکه‌های درون شهری با استفاده از شبیه‌سازی ترافیکی"، فصلنامه مهندسی حمل و نقل، دوره ۵، شماره٤، تابستان ۱۳۹۳، ص. ٤٧١–٤٨٤.

– ابرازی، روزبه، آقامحققی، مهناز و هاشمی تشکری، سید مهدی (۱۳۹۰) "روش‌های کنترل محدودیت سرعت به همراه مطالعه موردی بر روی بزرگراه صیاد شیرازی"، یازدهمین کنفرانس بین المللی مهندسی حمل‌ونقل و ترافیک.

– امین ناصری، محمد رضا و برادران، وحید (۱۳۸۹) "بررسی عوامل مؤثر بر زمان سفر در سیستم حمل و نقل عمومی و پیش‌بینی زمان سفر (موردکاوی: سیستم اتوبوسرانی شهر تهران)"، پژوهشنامه حمل و نقل، شماره ۳، ص ۲۱۹–۲۳۲.

- Rossi, R., Gastaldi, M., Gecchele, G., 2014, “Comparison of Clustering Methods for Road Group Identification in FHWA Traffic Monitoring Approach: Effects on AADT Estimates” Journal of Transportation Engineering, 140(7).

- Wen, F., Zhang, G., Sun, L., Wang, X. and Xu, X., 2019, “A hybrid temporal association rules mining method for traffic congestion prediction”, Computers & Industrial Engineering, 130, pp. 779-787.

- He, W., Lee T., and Wang E., 2013, “A New Method for Traffic Forecasting Based on the Data Mining Technology with Artificial Intelligent Algorithms”, Research

Journal of Applied Sciences, Engineering and Technology, 5(12), pp. 3417-3422.

- Park H.J., Kim, P.H., Marsico, M., Rasheed, N., 2014, "Data Mining Strategies for Real-Time Control in New York City", Procedia Computer Science, 32, pp.109 – 116.

- Gürbüza, F., Turna, F., 2018, Rule extraction for tram faults via data mining for safe transportation, Transportation Research Part A, 116, pp. 568-579.

- Tan, P.-N., Steinbach M., Kumar V., 2006, "Introduction to Data Mining". Pearson Addison-Wesley.

- Zamani, Z., Pourmand, M., Saraee, M.H., 2010, "Application of Data Mining in Traffic Management: Case of City of Isfahan", Procceding of International Conference on Electronic Computer Technology (ICECT), Kuala Lumpur, pp. 102 - 106.

- Moradpour, S., Long, S., 2019, Using combined multi-criteria decision-making and data mining methods for work zone safety: A case analysis, Case Studies on Transport Policy, 7, pp. 178-184.

- Caceres N., Romero L. M., Benitez F. 2012, "Estimating Traffic Flow Profiles According to a Relative Attractiveness Factor", In 15th Meeting of the EURO Working Group on Transportation, September 2012, Paris, Procedia - Social and Behavioral Sciences, 54, pp. 1115–1124.

- Bellemans, T., Schutter, B. D., Moor, B. D., 2000, "Data acquisition, interfacing and pre-processing of highway traffic data", Delft University of Technology, Faculty of Information Technology and Systems 1 (6). Website:http://www.dcsc.tudelft.nl