نشریه علمی پژوهشی مهندسی فناوری اطلاعات مکانی

سال سوم، شماره دوم، تابستان ۱۳۹۴

Vol.3, No.2, Summer 2015

۴۳-۵۹

پیش‌بینی و کنترل ترافیک با رویکردهای داده‌کاوی با استفاده از داده‌های GPS

زهرا مهدویان[1*]، علی‌اکبر نیک‌نفس[2]

۱- کارشناس ارشد مهندسی فناوری اطلاعات، دانشگاه تحصیلات تکمیلی صنعتی و فناوری پیشرفته کرمان، کرمان

۲- استادیار بخش مهندسی کامپیوتر، دانشگاه شهید باهنر کرمان،کرمان

تاریخ دریافت مقاله ۱۳۹۴/۰۴/۲۳ تاریخ پذیرش مقاله: ۱۳۹۴/۰۹/۲۹

چکیده

در دنیای امروز افزایش چالش‌های شهرنشینی و ترافیک سبب نیاز فوری به سیستم‌های کنترل ترافیک با حداکثر بهره‌وری شده است. هدر رفتن زمان و افزایش سوخت مصرفی و هم‌چنین آلودگی‌های هوا و صوتی سبب شده کنترل ترافیک به یکی از مهم‌ترین بحث‌های روز دنیا تبدیل شود. از جمله روش‌های موجود برای رسیدن به این هدف، پیش‌بینی مسیر و مقصد نهایی خودرو است. اگر مکان آینده خودروها را بتوان پیش‌بینی کرد، به‌راحتی می‌توان قادر به تخمین ازدحام ترافیک شد. پیش‌بینی مسیر براساس مسیرهای طی شده گذشته خودرو و در نظر گرفتن مشخصه‌هایی مانند مکان شروع، ساعت، روز، ماه، مدت زمان با استفاده از روش‌های داده‌کاوی و شبکه‌های عصبی مصنوعی امکان‌پذیر است. در این مقاله از داده‌های واقعی GPS، به‌دست‌آمده از خودروها به‌منظور انجام عملیات پیش‌بینی مسیر و مقصد نهایی بهره برده شده است. یکی از روش‌های پیشنهادی در این مقاله به‌دست آوردن پایگاه‌داده‌ای از مسیرهای دقیق طی‌شده خودروها به کمک نرم‌افزار ArcGIS است، که این پایگاه‌داده سبب بالا رفتن دقت پیش‌بینی مسیر خودرو شده است. در این مقاله یک پایگاه‌داده کلی شامل همه حالات شبکه جاده‌ای و یک پایگاه‌داده پیچیده‌تر شامل چهارراه‌ها و نقاط چالشی‌تر جاده ایجاد شده است. هم‌چنین به‌منظور عملیات پیش‌بینی، دو الگوریتم قوانین انجمنی و شبکه‌های عصبی مصنوعی مورد استفاده قرار گرفته شده است. نتایج به‌دست‌آمده نشان‌دهنده دقت بسیار خوب پیش‌بینی است. برطبق ارزیابی انجام‌شده، الگوریتم ANN در پایگاه‌داده کلی با دقت پیش‌بینی بالای ۹۶ درصد و الگوریتم GRI در پایگاه‌داده پیچیده‌تر با دقت پیش‌بینی بالای ۹۵ درصد، نتایج قابل قبولی به‌دست آورده‌اند. روش پیشنهادی و نتایج این پیش‌بینی می‌تواند به برنامه‌ریزی ترافیک و بهینه‌سازی حرکت خودروها کمک کند.

واژه‌های کلیدی: ترافیک، داده کاوی، پیش‌بینی، GPS.

* نویسنده مکاتبه کننده: کرمان، دانشگاه تحصیلات تکمیلی صنعتی فناوری پیشرفته کرمان، دانشکده برق و کامپیوتر

تلفن: ۰۹۱۶۳۶۸۷۱۷۵

Email: z.mahdavian@student.kgut.ac.ir

۱- مقدمه

با افزایش نرخ تولید و مهاجرت به مناطق شهری، از شبکه‌های جاده‌ای اغلب بیش از ظرفیت مورد نظر استفاده می‌شود. این باعث شده است بسیاری از شهرستان‌ها با به اجرا درآوردن اقدامات خاص مانند محدود کردن تردد وسایل نقلیه براساس شماره پلاک خودرو سعی در کنترل ترافیک کنند. درنتیجه نیاز فوری به درک و در حالت ایده‌آل پیش‌بینی ترافیک در شهرستان‌ها، برای کاهش تراکم ترافیک، هم‌چنین نیازهای زیست‌محیطی، اقتصادی و اجتماعی برای حمایت از یک آینده پایدار لازم است. در گذشته با نصب سنسورهای ترافیک در بخش‌های مختلف از شهر عملیات نظارت و آگاهی از ترافیک انجام می‌شد، که این معمولاً تنها در بزرگراه‌ها مورد استفاده قرار می‌گیرد[۱].

در چند سال گذشته، استفاده از سیستم‌های جهت‌یابی ماهواره‌ای از جمله گیرنده‌های GPS[1] - درون خودرویی در کشورهای توسعه‌یافته، در حال افزایش است. در همین زمان تلفن همراه و فناوری‌های بی‌سیم دارای پوشش گسترده جغرافیایی هستند که می‌توانند برای انتقال هم‌زمان اطلاعات خودرو به یک مرکز داده برای پردازش مورد استفاده قرار بگیرند. به‌عنوان مثال، سیستم‌های ردیابی خودرو از جمله سیستم رادیویی بسته عمومی[2]، مودم‌های نسل سوم (3G) یا دستگاه‌های PDA همگی می‌توانند داده‌های GPS خودرو را در زمان واقعی به هر کامپیوتر منتقل کنند، که بعد جدیدی در زمینه تحقیقات سیستم‌های حمل‌ونقل هوشمند به‌خصوص در تشخیص حادثه و مدیریت ترافیک ایجاد کرده است[۲].

سیستم‌های حمل‌ونقل هوشمند[3] برنامه‌های کاربردی پیشرفته‌ای هستند که هدف این سیستم‌ها ارائه خدمات نوآورانه‌ای در زمینه حمل‌ونقل و مدیریت ترافیک است، که کاربران را قادر می‌سازد از شبکه‌های حمل‌ونقل هماهنگ‌تر، امن‌تر و دقیق‌تر بهره ببرند. هم‌چنین GPS در زمینه‌های مختلف مثل ساخت‌وساز نقشه‌های اتوماتیک[۳] تشخیص نقاط ازدحام[۴] محاسبات شهری[۶،۵] و دیگر زمینه‌ها مورد استفاده قرار گرفته است.

ازدحام ممکن است با دادن اطلاعات دقیق و به‌موقع به رانندگان کاهش یابد. با این اطلاعات، رانندگان می‌توانند با جایگزین کردن مسیرهای دیگر یا تغییر زمان حرکت خود از ترافیک جلوگیری کنند. بر این اساس اطلاعات ترافیکی یک عنصر کلیدی برای آگاه کردن رانندگان به منظور تصمیم‌گیری در مورد مسیر حرکتی خود و درنتیجه پیش‌گیری از ازدحام در نظر گرفته می‌شود. داده‌های به‌دست‌آمده از GPS، اطلاعات کافی را برای تشخیص الگوهای ترافیکی، مانند متوسط سرعت در یک جاده مشخص‌شده و رفتارهای غیرمعمول خودرو مانند تغییر جهت حرکت، فراهم می‌آورد. GPS قابل قبول‌ترین، دقیق‌ترین، اقتصادی‌ترین و انعطاف‌پذیرترین تکنولوژی مکان‌یابی است که به منظور به دست آوردن داده‌های زمانی سفر، استفاده می‌شود[۳].

از جمله روش‌های به دست آوردن اطلاعات ترافیکی، پیش‌بینی مسیر و مقصد نهایی خودرو است. پیش‌بینی مسیر خودرو براساس مسیرهای طی‌شده توسط خودرو در گذشته، از مباحث ویژه در سیستم‌های حمل‌ونقل هوشمند محسوب می‌شود.

از جمله مسئله مهمی که در این مقاله به آن پرداخته شده است، افزایش دقت پیش‌بینی مسیر خودرو به منظور درک بهتر از جریان ترافیک و ازدحام است، که این مشکل با ساخت یک مدل از جریان ترافیک در شهر براساس مشاهدات تاریخی قابل حل است.

مدل ایجادشده در این مقاله براساس الگوریتم‌های پیش‌بینی در داده‌کاوی است که لازمه این مدل داشتن یک پایگاه‌داده دقیق از مسیرهای طی‌شده

[1] Global Positioning System
[2] General Packet Radio System (GPRS)
[3] Intelligent transportation systems (ITS)

توسط خودرو است. به این دلیل برای ایجاد یک پایگاه‌داده از مسیرهای طی‌شده توسط خودرو، از نرم‌افزار ArcGIS به کمک داده‌های GPS به‌دست‌آمده از خودروها بهره برده شده است. این پایگاه‌داده به‌منظور قطعه‌بندی دقیق جاده‌ها و به‌دست آوردن مسیرهای طی‌شده توسط خودرو ایجاد شده و در روند روش پیشنهادی سبب افزایش دقت پیش‌بینی مسیر شده است.

در این مقاله با بهره‌گیری از روش قوانین انجمنی و شبکه‌های عصبی مصنوعی، مسیر خودروها و مقصد مورد نظر خودرو پیش‌بینی شده است. قوانین انجمنی قادر به شرح الگوی رفتاری رخدادهایی است که به‌ترتیب متوالی اتفاق افتاده‌اند. در طول یک جاده، این رخدادها، به‌عنوان یک لیست مرتب‌شده از قطعه‌های جاده طی‌شده توسط خودرو در طول سفر خود از نقطه شروع تا مقصد مورد نظر است. قوانین انجمنی از روش داده‌کاوی به‌منظور پیدا کردن ارتباط‌هایی میان رخدادهای به ترتیب زمانی و در ارتباط با یکدیگر، بهره می‌گیرد. شبکه‌های عصبی مصنوعی مدل پیش‌بینی را طی یک فرآیند یادگیری از داده‌های موجود فرا می‌گیرند.

۲- کارهای مرتبط

یانگ در سال ۲۰۰۵، با استفاده از روش فیلتر کالمن و روش برآورد به پیش‌بینی زمان سفر متمرکز شده است. یانگ به سه دلیل زیر، زمان سفر را معیار عملکرد قرار داده است. ۱. اندازه‌گیری زمان رایج‌ترین روشی است که کاربران می‌توانند کیفیت سفر خود را اندازه‌گیری کنند. ۲. متغیر زمان به‌طور مستقیم قابل اندازه‌گیری است. ۳. یک اقدام ساده برای نظارت بر ترافیک است. در این تحقیق یانگ مطالعه خود را روی مدت زمان خارج شدن از یک منطقه که به‌طور پیوسته دارای ازدحام است، انجام داده است[۷].

شهیب کامران [۲] در سال ۲۰۰۷، یک روش چندسطحی به‌منظور تشخیص ترافیک رخ داده‌شده به‌علت تصادف در یک جاده اصلی ارائه کرده است. در این روش با استفاده از داده‌های زمان واقعی GPS به‌دست‌آمده از خودرو، الگوریتم‌هایی برای تشخیص الگوهای ترافیکی غیرمعمول و رفتار خودرو در بخش‌های مختلف جاده ارائه شده‌است. بر طبق این مقاله تشخیص تصادف شامل دو فاز است: ۱. شناسایی بخش‌هایی از جاده که در آن الگوی ترافیکی غیرطبیعی مشاهده شده است و افزایش تقسیم‌بندی بخش‌های غیرعادی به بخش‌های کوچک‌تر به‌منظور جدا کردن منطقه بالقوه حادثه از دیگر مناطق. ۲. انجام سلسله مراتبی تجزیه و تحلیل داده‌های GPS خودروها براساس قوانین از پیش تعریف‌شده به‌منظور تشخیص هرگونه رفتار غیرعادی در بخش‌های غیرعادی مشخص‌شده در فاز یک است. مزیت اصلی این روش در تجزیه و تحلیل داده‌های خودروهایی است که در مناطق شناسایی‌شده نهفته شده است که با این روش از پردازش داده‌های گسترده در دیگر مناطق جلوگیری می‌شود و این امر سبب عملکرد بهتر شده است.

تئوری چیدمان فضا[1] یک روش شناخته‌شده تجزیه و تحلیل است که بر اساس استفاده از روش‌های کامپیوتری به تحلیل ساختاربندی شهری می‌پردازد. این روش می‌تواند به دو مؤلفه تقسیم شود؛ تحلیل شبکه فضاهای عمومی انتخاب‌شده و ارائه یک نقشه و گراف که ارتباط و اتصالات این فضاها را شرح می‌دهد[۸]. خالصیان و همکاران[۹] در سال ۲۰۰۹، با استفاده از تئوری چیدمان فضا، مدل‌سازی ترافیک پویا و قابلیت‌های GIS، درجه اهمیت تقاطع خیابان‌های درون‌شهری را اندازه‌گیری کردند. این اندازه‌گیری کاربرد مهمی در سیستم‌های مدیریت ترافیک پیشرفته[2] دارد. سیستم‌های مدیریت ترافیک پیشرفته یکی از مؤلفه‌های سیستم‌های انتقال هوشمند

[1] Space Syntax
[2] Advanced traffic management systems (ATMS)

است، که به‌منظور پیش‌بینی حجم ترافیکی و فراهم آوردن اطلاعات ترافیکی برای بزرگراه‌ها مورد استفاده قرار می‌گیرد.

جیانگ و لیو بر روی یک توپولوژی به منظور تجزیه و تحلیل پیش‌بینی جریان ترافیک در یک محیط شهری تمرکز کرده‌اند. در این روش براساس تئوری چیدمان فضا، هر خیابان به صورت خطوط محوری نشان داده شده است، که به اصطلاح یک نقشه محوری[1] نامیده می‌شود. در این مقاله توپولوی خیابان–خیابان[2] ایجاد شده است که نتایج به‌دست‌آمده نشان‌دهنده بهبود پیش‌بینی جریان ترافیک در یک محیط شهری است [۱۰].

لی یانینگ و مایک مک دونالد یک الگوریتم براساس ماشین کاوش‌گر برای تشخیص تصادف در بزرگراه‌ها ارائه کرده‌اند. الگوریتم براساس مدل تجزیه و تحلیل دو متغیره[3] با استفاده از دو متغیر، متوسط زمان سفر ماشین کاوش‌گر و اختلاف زمان سفر بین فواصل زمانی هم‌جوار ارائه شده است[۱۱]. اسوکاف و همکاران یک فرآیند که مدل‌های پیش‌بینی داده‌کاوی را به منظور پیش‌بینی ترافیک ترکیب می‌کند، برای استفاده در سیستم‌های اطلاعات ترافیکی ارائه داده‌اند. در مقاله مذکور از مدل خوشه‌بندی برای کشف ارتباط بین سرعت- جریان- ظرفیت در ترافیک استفاده شده است. نتاج نشان می‌دهد که مدل ارائه‌شده قادر به پیش‌بینی پارامترهای ترافیکی با دقت بالاست[۱۲].

اوزبای و همکاران یک روش جدید با استفاده از داده کاوی برای شناسایی خودکار حوادث بزرگراه ارائه کرده‌اند. مدل به‌عنوان یک جزء زمان واقعی از سیستم کنترل تطبیقی ترافیک برای کنترل سیگنال است. الگوریتم هرگاه یک حادثه را احساس کرد، یک گزارش برای مدیر بهینه‌سازی ارسال می‌کند که با استفاده از اطلاعات به‌دست‌آمده استراتژی مناسب منظور شود. این روش به صورت آزمونی آن‌لاین برای اثبات عملکرد شبیه‌سازی شده است. از آن‌جا که جریان ترافیک معمولاً اتفاقی و نامطمئن مشخص می‌شود، منطق فازی به‌خوبی برای مدل‌سازی و کنترل چنین مشکلاتی شناخته شده است. برنامه‌های کاربردی منطق فازی به منظور کنترل سیگنال ترافیک ساخته شده است[۱۳]. همچنین الگوریتم‌ها و کنترل کننده‌های منطقی مختلفی برای جریان عادی ترافیک توسعه داده‌اند [۱۴،۱۵].

۳- مروری بر مفاهیم پایه

در این بخش توضیحات مختصری در مورد مفاهیم پایه و تعاریف مرتبط با موضوع مورد بحث آورده شده است.

۳-۱- تعریف داده‌کاوی و روش‌های آن

داده‌کاوی عبارت است از فرآیند استخراج اطلاعات معتبر، از پیش ناشناخته، قابل فهم و قابل اعتماد از پایگاه‌داده‌های بزرگ به منظور کشف الگوهای پنهان[۱۶،۱۷]. در ادامه توضیحات مختصری در مورد عملیات داده‌کاوی از جمله دسته‌بندی، قوانین انجمنی و شبکه‌های عصبی مصنوعی ارائه می‌شود.

- دسته‌بندی: دسته‌بندی فرآیندی است که مجموعه داده‌ها را به قسمت‌های مشخص تقسیم می‌کند. برای مثال خودروهای یک جاده براساس سرعتشان به دو گروه با سرعت بالا و سرعت پایین تقسیم می‌شوند. دسته‌بندی را می‌توان به‌عنوان یک فرآیند دو مرحله‌ای در نظر گرفت. در مرحله اول، یک مدل دسته‌بندی با توجه به مجموعه داده‌های آموزشی ساخته می‌شود، چنین مدلی می‌تواند به فراهم کردن یک درک بهتر از داده‌های گسترده کمک کند. سپس قوانین اگر- آن‌گاه جهت پیش‌بینی، مورد استفاده قرار می‌گیرند[۱۸].
- قوانین انجمنی: قوانین انجمنی به‌منظور استخراج اطلاعاتی در مورد وقوع رخدادهای معین براساس

[1] Axial Line
[2] Street- Street Topology
[3] Bivariate Analysis Model(BEAM)

حداقل مقدار پشتیبانی تعیین‌شده در ضوابط پایه‌ای برای چگونگی انجام فرآیند اکتشاف در داده‌کاوی، استفاده می‌شود[۲۰،۱۹].

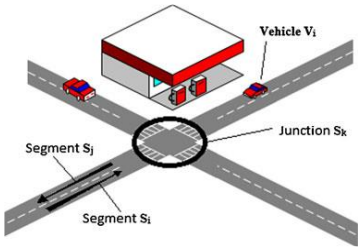
درواقع قوانین انجمنی به مجموعه‌ای از قوانین گفته می‌شود که بیان‌گر رخ دادن دو یا چند واقعه به همراه هم است. به عبارت دیگر این قوانین وقوع هم‌زمان رخدادها را در قالب اگر – آن‌گاه نمایش می‌دهند. برای ارزیابی قدرت یک قانون پارامترهای پشتیبانی[1] و اطمینان[2] تعریف شده است.

- شبکه‌های عصبی مصنوعی: شبکه‌های عصبی مصنوعی یکی از موضوعاتی است که در مدل‌سازی، پیش‌بینی، رده‌بندی و... با حجم داده بسیار زیاد و داده‌کاوی استفاده می‌شود. اساس این روش برپایه مغز انسان است که با توجه به نوع کارکرد مغز دریافت اطلاعات، پردازش آن‌ها و درنهایت تولید یک خروجی پایه‌ریزی شده است. یک ساختار شبکه‌ای از تعدادی عناصر مرتبط به هم به نام نرون تشکیل می‌شود، که هر نرون دارای ورودی‌ها و خروجی‌هایی است و یک عمل نسبتاً ساده و محلی را انجام می‌دهد. شبکه‌های عصبی مصنوعی عموماً عملکرد خود را طی یک پروسه یادگیری فرامی‌گیرند. شبکه‌های عصبی کاربردهای عمده‌ای در تشخیص الگو، گروه‌بندی، پیش‌بینی یا برون‌یابی و... دارند[۲۱].

۳-۲- تعاریف الگوهای حرکتی

به‌منظور بهره‌گیری از قوانین انجمنی در پیش‌بینی مسیر خودرو نیاز است تعاریفی متناسب با الگوهای حرکتی خودرو صورت بگیرد. در این بخش شرح تعاریف الگوهای حرکتی ماشین‌ها فراهم آورده شده، که برگرفته از تعاریف الگوهای حرکتی ارائه‌شده در مراجع [۲۳،۲۲] است.

تعریف ۱: یک مجموعه $S = \{S1, S2, \dots, Si\}$ نشان‌دهنده قطعه‌های موجود در یک نقشه است. یک قطعه به‌عنوان یک لبه از جاده در نظر گرفته می‌شود که دو گره را به یکدیگر وصل می‌کند. یک گره به‌عنوان یک نقطه اتصال در جاده است که می‌تواند یک چهارراه، انتهای جاده، محل خروج و... در نظر گرفته شود. همان‌طور که در شکل ۱ نشان داده شده است، دو مسیر رفت و برگشت، دو قطعه جدا از هم در نظر گرفته می‌شوند.



شکل۱: قطعه‌بندی و تقاطع جاده[۲۲]

تعریف ۲: مجموعه $M = [S1, S2, \dots, Sn]$ یک الگوی حرکتی است که نشان‌دهنده قطعه‌های طی‌شده توسط یک ماشین در طول مسیر خود در یک منطقه جغرافیایی مشخص در نقشه مورد نظر است.

تعریف ۳: مجموعه $V = \{V1, V2, \dots, Vn\}$ نشان دهنده مجموعه ماشین‌های گذرنده در یک منطقه جغرافیایی در یک فاصله زمانی معین است. پایگاه‌داده حرکت ماشین‌ها که در جدول ۱ نشان داده شده است، نشان‌دهنده مجموعه الگوی حرکتی هر ماشین در طول مسیر خود است. هر ماشین به‌عنوان یک درایه در نظر گرفته می‌شود.

جدول۱ : پایگاه‌داده حرکت خودروها

الگوهای حرکتی	شناسه خودرو
$[S_1,S_3,S_5,S_9]$	V_1
$[S_3,S_8,S_{12},S_{13}]$	V_2
...	.....
$[S_1,S_2,\dots,S_n]$	V_n

تعریف ۴: یک الگوی $M = [S1, S2, \dots, Sn]$ به‌عنوان یک زیرالگو M’ در نظر گرفته می‌شود، اگر عناصر آن

[1] Support
[2] Confidence

به همان ترتیب که در M' هستند، استخراج شده باشند. برای نمونه، الگوی [S3, S4, S6] یک زیرالگو برای الگوی [S1, S3, S4, S6] است، اما یک زیرالگو برای [S1, S3, S4, S5, S6] نیست.

تعریف ۵: پشتیبان الگوی حرکتی M، support (M)، به‌عنوان تعداد الگوها در پایگاه‌داده حرکت است که M به‌عنوان زیرالگو برای آن‌ها تعریف می‌شود.

تعریف ۶: اگر M = [S1, S2, ..Sn] یک الگوی حرکتی است، قانون حرکت R با دلالت بر M1 => M2 تعریف می‌شود. جایی که M1 و M2 زیرالگوهای M هستند و الگوهای (قطعه‌ها) بین M1 و M2 وجود نداشته باشد، اعتماد الگوی R، conf(R)، توسط الگوی زیر تعریف می‌شود.

$$Conf(R) = \frac{Support(M)}{Support(M_1)}$$ رابطه (۱)

تولید الگوی حرکتی نیازمند جمع‌آوری رفتار حرکتی ماشین‌هاست و قوانین تولیدشده از الگوهای حاضر در پایگاه‌داده حرکتی، ایجاد می‌شوند. برای نمونه، اگر M= [S1, S2, S3, S4] یک الگو باشد. مجموعه قوانین ممکن استخراج‌شده به صورت زیر هستند.

{[S1] ⇒ [S2, S3, S4], [S1, S2] ⇒ [S3, S4], [S1, S2, S3] ⇒ [S4]} رابطه (۲)

۳-۲- تعاریف GPS

سیستم موقعیت‌یاب جهانی (GPS) سرویس‌های زمانی و مکانی و جهت‌یابی را برای کاربران فراهم می‌کند. این سیستم شامل قسمت فضایی، قسمت کنترل و قسمت کاربری است.

۱. قسمت فضایی شامل ۲۴ ماهواره در صورت فلکی است که موقعیت ماهواره‌ای GPS و زمان کنونی آن را به‌صورت سیگنال‌های یک‌طرفه ارسال می‌کند.

۲. قسمت کنترل شامل ایستگاه‌های نظارت و کنترل در سراسر جهان است که مسئولیت حفظ ماهواره‌ها در مدارهای مناسب خود را به‌عهده دارد.

۳. قسمت کاربری شامل ابزارهای دریافت‌کننده GPS است که سیگنال‌ها را از ماهواره‌های GPS دریافت می‌کند و اطلاعات انتقال داده‌شده برای محاسبه زمان و موقعیت مکانی کاربر استفاده می‌شود[۴].

۳-۲-۱- برنامه‌های کاربردی GPS

- زمان‌بندی: سیستم موقعیت‌یاب جهانی، زمان دقیق و واقعی ثبت اطلاعات را نیز علاوه بر مشخص کردن طول، عرض و ارتفاع جغرافیایی فراهم می‌کند. هر ماهواره GPS شامل چندین ساعت اتوماتیک است که زمان دقیق را برای سیگنال‌های GPS فراهم می‌کند. زمان یکی از بااهمیت‌ترین معیارها برای انجام فعالیت‌های نظیر پیش‌بینی مسیر خودرو است.
- تحلیل ترافیک جاده‌ها: تخمین‌ها نشان می‌دهد که تأخیرهای ناشی از ازدحام‌های بزرگراه‌ها، خیابان‌ها و سیستم‌های حمل‌ونقل در جهان سالانه هزار میلیون دلار هزینه در برداشته است. از دیگر معایب ازدحام و ترافیک ضررهای احتمالی شامل صدمه دیدن اشخاص، افزایش آلودگی هوا و مصرف سوخت است. با ترکیب تکنولوژی مکانیGPS با سیستم‌هایی که قادر به نمایش اطلاعات جغرافیایی یا انتقال خودکار داده به کامپیوتر هستند، بعد جدیدی در انتقال ایمن فراهم می‌کنند. هم‌چنین با استفاده از سیستم اطلاعات جغرافیایی GIS[1]، ذخیره، تجزیه و نمایش اطلاعات فراهم‌شده از GPS امکان‌پذیر شده است. امروزه GIS برای مانیتور کردن موقعیت خودرو به منظور ایجاد استراتژی‌های مؤثر مانند زمان‌بندی دقیق سفر و ردیابی مسیر مورد استفاده قرار می‌گیرد[۴].

[1] Geographic Information Systems (GIS)

۴- شرح مسئله و روش پیشنهادی

راه‌حلی که ما برای بحث کنترل و برنامه‌ریزی ترافیک شهری در نظر گرفته‌ایم، پیش‌بینی مسیر خودروهاست. معمولاً رانندگان از مسیرهای مشخصی برای رسیدن به مقصد عبور می‌کنند، به همین منظور با استفاده از داده‌های تاریخی به‌دست‌آمده از حرکت خودروها که شامل مشخصه‌هایی از جمله مکان شروع سفر، زمان انجام یک سفر، روز، ماه، مدت زمان و خیابان‌های طی‌شده است، پیش‌بینی مسیر بعدی و مقصد خودرو امکان‌پذیر شده است. بنابراین با تحلیل رفتارهای گذشته رانندگان توسط روش‌های داده‌کاوی و الگوریتم‌های پیش‌بینی، می‌توان مسیر حرکت بعدی رانندگان را برای مقصد مشخصی پیش‌بینی کرد.

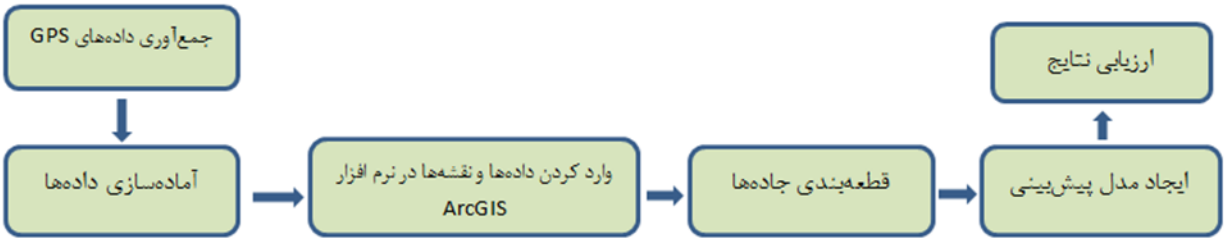
نتایج حاصل از پیش‌بینی الگوریتم‌های شبکه‌های عصبی مصنوعی و روش‌های داده‌کاوی تعبیه‌شده در ابزارهای درون‌خودرویی هر خودرو، با بهره‌گیری از شبکه‌های بین‌خودرویی[1] و برقراری ارتباط با واحدهای کنارجاده[2] به‌صورت سلسه مراتبی به مرکز کنترل ترافیک ارسال می‌شود. به این ترتیب اطلاعات مربوط به مقصد همه خودروها در دسترس مرکز کنترل ترافیک قرار می‌گیرد و امکان محاسبه حجم ترافیکی هر منطقه قابل انجام است. سپس نتایج پیش‌بینی حجم ترافیکی هر منطقه توسط واحدهای کناره جاده و شبکه‌های بین خودرویی از مرکز کنترل ترافیک به هر خودرو اطلاع داده می‌شود.

بخشی از اهداف در این مقاله به دست آوردن یک پایگاه‌داده از مسیرهای طی‌شده توسط خودرو است. پایگاه‌داده به‌دست‌آمده به میزان قابل توجهی در دقت پیش‌بینی مسیر خودرو مؤثر است. برای به دست آوردن این پایگاه‌داده از نرم‌افزار ArcGIS استفاده شده است، به کمک این نرم‌افزار عملیات قطعه‌بندی جاده‌ها صورت گرفته است.

به منظور انجام عملیات داده‌کاوی از الگوریتم شبکه‌های عصبی مصنوعی و دو مدل از قوانین انجمنی شامل قوانین باهم‌آیی و ساختار قواعد پیوند[3] استفاده شده است. قوانین باهم‌آیی زیرمجموعه‌ای از قوانین انجمنی است، که در این الگوها ترتیب به وقوع پیوستن رخدادها مهم است، برای مثال رخداد S1⇒S2 با رخداد S2⇒S1 متفاوت است. به‌طور کلی داده‌های ورودی در الگوریتم باهم‌آیی شامل لیستی از تراکنش‌ها به همراه زمان رخداد آن‌هاست، که در این‌جا تراکنش‌ها شامل قطعه‌های طی‌شده متوالی جاده توسط خودرو است. هم‌چنین الگوریتم GRI به صورت دو عبارت مقدم و تالی است. این الگوریتم برای رسیدگی به داده‌های حجیم و به‌صورت عددی بسیار کارآمد است.

عملیات مربوط به داده‌کاوی در نرم‌افزار داده‌کاوی Clementine12.0 انجام شده است. هدف اصلی از انجام عملیات داده‌کاوی پیدا کردن قوانینی از حرکت خودروهاست که با ارائه این کار پیش‌بینی مسیر و مقصد آینده خودرو امکان‌پذیر است. درنتیجه حجم ترافیکی قطعه‌های جاده قابل پیش‌بینی و عملیات برنامه‌ریزی و کنترل ترافیک به‌راحتی قابل انجام است. درنهایت عملیات ارزیابی با مقایسه نتایج پیش‌بینی و مقادیر واقعی مورد سنجش قرار گرفته شده که نتایج حاصل در بخش ۵ آورده شده است. مراحل انجام کار در این مقاله در شکل ۲ آورده شده است.

[1] Vehicular Ad hoc Network (VANET)
[2] Road Side Unit (RSU)
[3] GRI : Generalized Rule Induction



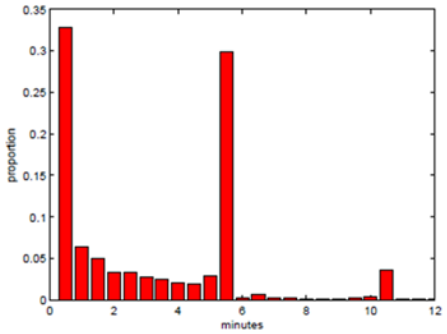
شکل۲ : مراحل انجام کار

۴-۱- پایگاه‌داده

برای ارائه اطلاعات ترافیکی چندین همکاری مشترک بین شرکت‌های خصوصی و سازمان‌های دولتی برای جمع‌آوری اطلاعات از ترافیک‌های جاده‌های اصلی و گزارش وضعیت ترافیک به مردم انجام شده است. این شرکت‌ها اطلاعات ترافیک را از طریق حلقه‌های جمع‌آوری، دوربین و برچسب‌های عوارضی الکترونیکی که بسیار مؤثر هستند، به‌دست می‌آورند. اما مشکل اصلی هزینه بسیار بالای این سیستم هاست[۲۴].

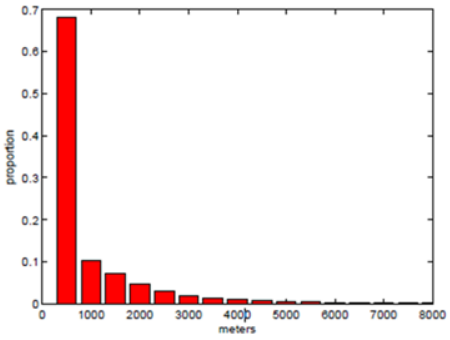
یکی از روش‌های کاربردی به‌منظور کنترل ترافیک استفاده از داده‌های آن‌لاین در مبحث جریان داده[۱] است. یک جریان داده درواقع یک توالی از داده‌هایی به شکل $(x_1,\ldots,x_n)$ است، که هر کدام از x_iها یک خصوصیت دارای مقدار هستند[۲۵]. یک صورت مسئله ترافیک باید اطلاعات آن‌لاین را از یک شبکه حس‌گرهای بی‌سیم یا GPS دریافت کند و سرورهایی به‌صورت هم‌زمان با ثابت زمانی کم داده‌کاوی را انجام دهند و در مقطع زمانی با تأخیر قابل قبول نتایج را به اطلاع خودرو برسانند.

در این مقاله به علت در دسترس نبودن داده‌های آن‌لاین، از تکنیک GPS برای جمع‌آوری داده به صورت آف‌لاین، به منظور تشخیص موقعیت مکانی خودروها و محاسبه حجم ترافیکی خیابان‌های مختلف استفاده شده است. پایگاه‌داده موجود شامل ۱۵میلیون رکورد از نقاط مکانی GPS، نصب‌شده روی ۱۰۳۵۲ تاکسی در طول دوره دوم تا هشتم فوریه سال ۲۰۰۸ میلادی شهر پکن در کشور چین است. ابزار GPS نصب‌شده روی خودروها به صورت دوره‌های زمانی مرتب به جمع‌آوری اطلاعات مکانی خودرو پرداخته‌اند. بر طبق نمودار ۱ بازه زمانی بیشتر نقاط ثبت‌شده بین سه تا پنج ثانیه است. هم‌چنین بازه مسافت طی‌شده بین دو نقطه ثبت‌شده بر طبق نمودار ۲ تقریباً بین ۴۰۰ تا ۷۰۰ متر است[۲۶]. جدول ۲ نمونه‌ای از رکوردهای ثبت‌شده از نقاط مکانی GPS در پایگاه‌داده موجود است. به‌طور نمونه رکورد نشان داده‌شده در جدول بیان می‌کند که خودرو با شناسه ۱۱۳۱ در روز دوم فوریه در ساعت ۱۳:۳۴:۰۹ در موقعیت مکانی با طول و عرض جغرافیایی نشان داده‌شده، بوده است.



نمودار۱: توزیع بازه زمانی بین دو نقطه متوالی[۲۶]

[۱] Data Stream



نمودار۲ : توزیع بازه مسافتی بین دو نقطه متوالی [۲۶]

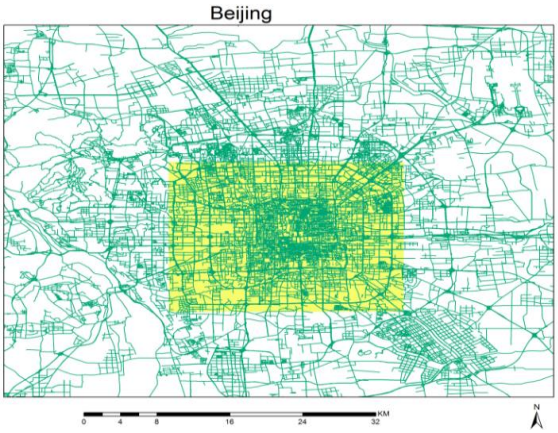
جدول۲ : نمونه‌ای از رکورد نقاط ثبت‌شده GPS در پایگاه‌داده موجود[۲۶]

TaxiID	Date	Time	Longitude	Latitude
1131	2008-02-02	13:34:09	116.51172	39.92123

۵- پیاده‌سازی و تحلیل نتایج

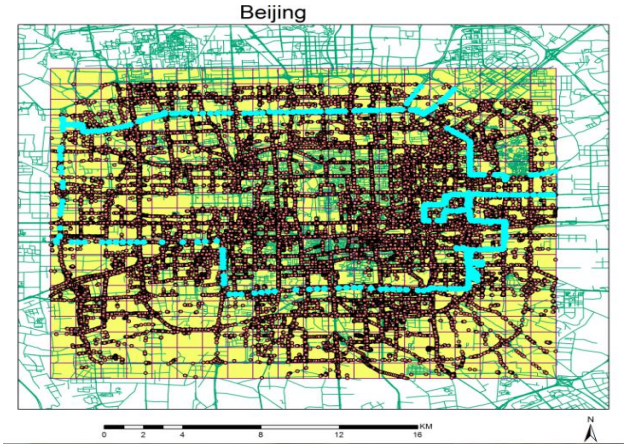
۵-۱- ایجاد پایگاه‌داده مسیر

همان‌طور که گفته شد، یکی از نیازهای اصلی برای ایجاد مدل پیش‌بینی مسیر، یک پایگاه‌داده از مسیرهای طی‌شده توسط خودرو است. به منظور قطعه‌بندی جاده و به دست آوردن مسیرهای طی‌شده توسط خودرو از نرم‌افزار ArcGIS بهره گرفته شده‌است. نقشه شهر پکن از کشور چین در شکل ۳ نشان داده شده است؛ محدوده مشخص‌شده در نقشه، منطقه مورد بررسی قرار گرفته‌شده در این مقاله است.



شکل۳: نقشه شهر پکن [۲۷]

داده‌های به‌دست‌آمده از GPS به صورت طول و عرض جغرافیایی مکان خودرو ذخیره شده‌اند. با وارد کردن این داده‌ها در نرم‌افزار، مسیر طی‌شده توسط خودرو به‌راحتی قابل مشاهده است. مسیر طی‌شده توسط یک خودرو در زمان و تاریخ معین در شکل ۴ نشان داده شده است.

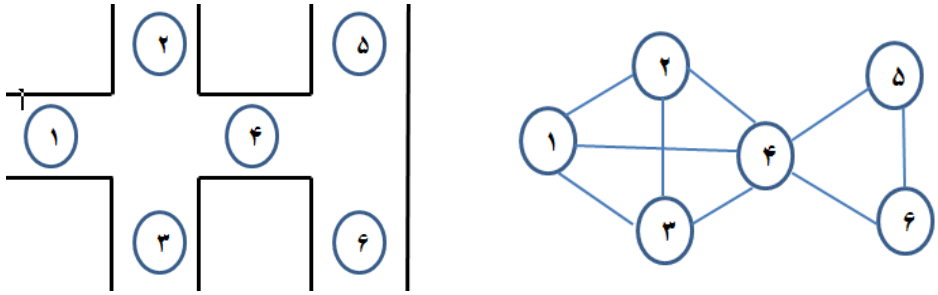


شکل۴: مسیر طی‌شده توسط خودرو در روز و زمان مشخص

به‌منظور ایجاد پایگاه‌داده مسیر از یک توپولوژی شبکه‌ای بهره گرفته شده است. که این روش گراف شبکه جاده‌ای[1] نامیده می‌شود.

نحوه قطعه‌بندی در این روش به این صورت است که هر تقاطع در جاده اعم از چهارراه، سه‌راه و هر عاملی در جاده که موجب تغییر جهت حرکت خودرو شود، به‌عنوان یک گره و جاده‌های بین هر دو گره یک یال در نظر گرفته شده است. درواقع یک گراف از شبکه جاده‌ای ایجاد می‌شود. همان‌طور که در شکل ۵ نشان داده شده است، قسمت شکل۵ (الف) نشان‌دهنده یک شبکه جاده‌ای است که شامل شش جاده است. گراف ایجادشده برای این شبکه در شکل۵ (ب) نشان داده شده است. این گراف شامل شش نود است که هر نود معرف یک تقاطع و هم‌چنین شامل ۹ یال است که معرف جاده‌های ارتباطی بین نودها است.

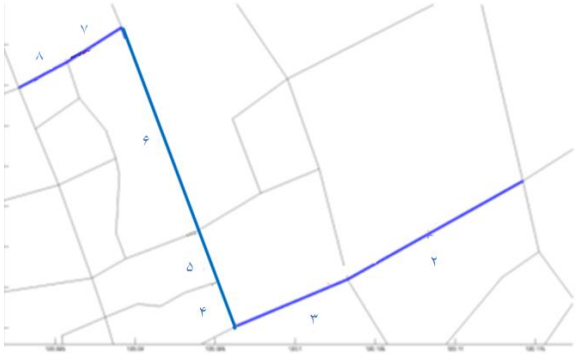
[1]Graph Road Network



(الف) (ب)

شکل۵: مثالی از گراف ارائه‌شده برای شبکه جاده‌ای. الف) یک شبکه با شش جاده ب) گراف شبکه جاده‌ای با شش گره و ۹ یال

سپس به هر یال و نود یک شناسه منحصربه‌فرد تعلق می‌گیرد. با عبور خودرو از هر یال شناسه آن یال توسط خودرو دریافت می‌شود. برای نمونه شکل۶ مسیر طی‌شده توسط خودرو در یک شبکه جاده‌ای را که به یک گراف تبدیل شده، نشان می‌دهد



شکل۶: مسیر طی‌شده توسط خودرو در یک شبکه جاده‌ای

. به این صورت مسیر طی‌شده یک خودرو به‌راحتی قابل دست‌یابی است. هم‌چنین علاوه بر شماره شناسه هر یال، تاریخ و زمان حرکت نیز ذخیره می‌شود. به این ترتیب داده‌های دقیق از حرکت خودروها متناسب با زمان و روز حرکت خودرو به دست می‌آید، سپس این داده‌ها در یک پایگاه‌داده مسیر مانند جدول ذخیره می‌شود. در این پایگاه‌داده مسیر، به ازای هر خودرو یک شماره شناسه (ID) و هم‌چنین تاریخ حرکت خودرو و بازه‌های زمانی مسیر طی‌شده توسط خودرو به‌عنوان ویژگی‌های دیگر پایگاه داده در نظر گرفته شده است.

جدول۳: نمونه‌ای از پایگاه‌داده استخراج‌شده در نرم‌افزار ArcGIS

ID	Date	StartTime	EndTime	Seg1	Seg2	Seg3	Seg4	Seg5	Seg6	Seg7	...	Seg15	...	Seg19	Seg20
366	02-Feb-2008 00:00:00	14:30:37	15:30:45	17803	17796	37460	37495	37497	37506	37508	...	15534	...	15535	15533
...	.....	...	...	...	...	...	...	...	...	...	...	...	...	...	...
366	02-Feb-2008 00:00:00	18:36:37	19:30:56	15536	18556	3047	15639	4027	38808	33373	...	23529	...	23523	23521
...	.....	...	...	...	...	...	...	...	...	...	...	...	...	...	...
456	02-Feb-2008 00:00:00	19:30:16	19:40:48	1357	39089	26836	1371	23649	23672	27362	...	31837	...	14853	38533
...	.....	...	...	...	...	...	...	...	...	...	...	...	...	...	...
456	02-Feb-2008 00:00:00	23:30:15	24:00:00	18942	197	18914	2390	18957	2384	2387	...	2533	...	23390	2183
...	.....	...	...	...	...	...	...	...	...	...	...	...	...	...	...
1131	02-Feb-2008 00:00:00	14:14:38	15:26:34	2303	14204	12648	31458	2301	2300	95	...	28724	...	26709	22360
...	.....	...	...	...	...	...	...	...	...	...	...	...	...	...	...
1131	03-Feb-2008 00:00:00	17:10:11	18:24:45	23077	4456	36316	23060	31937	4452	31904	...	31906	...	31944	21022
...	.....	...	...	...	...	...	...	...	...	...	...	...	...	...	...
1131	08-Feb-2008 00:00:00	12:20:19	13:30:34	31365	22717	8053	15401	15408	15407	10936	...	5107	...	27650	40105

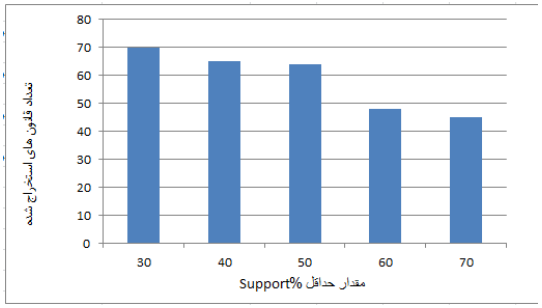
۵-۲- تحلیل نتایج

به منظور انجام عملیات پیش‌بینی از دو الگوریتم شبکه‌های عصبی مصنوعی و دو مدل قوانین انجمنی شامل الگوریتم باهم‌آیی و الگوریتم GRI در داده‌کاوی استفاده شده است. هم‌چنین برای مقایسه میزان دقت پیش‌بینی روش‌ها از دو پایگاه‌داده متفاوت استفاده شده است. پایگاه‌داده اول شامل همه حالات ممکن در شبکه جاده‌ای است که پایگاه‌داده کلی نامیده شده است. پایگاه‌داده دوم شامل چهارراه‌ها و حالت‌های چالشی‌تر جاده است که دارای پیچیدگی بالایی است.

ارزیابی نتایج حاصل به این صورت انجام گرفته است که در همه مدل‌های ایجادشده اعم از شبکه‌های عصبی مصنوعی و قوانین انجمنی داده‌های ورودی به دو دسته آموزش و آزمایش توسط نود پارتیشن[1] در نرم‌افزار کلمنتاین تقسیم شده‌اند. پایگاه داده مسیر استفاده‌شده در مدل‌ها شامل ۷۲۰۰ مسیر طی‌شده توسط خودروهای مختلف در روزها و زمان‌های متفاوت است. حدود ۸۰ درصد داده‌های ورودی به‌منظور آموزش الگوریتم‌ها در نظر گفته شده است، هم‌چنین حدود ۲۰ درصد داده‌های ورودی به‌عنوان داده‌های آزمایش به‌منظور ارزیابی دقت پیش‌بینی الگوریتم‌ها قرار داده شده است. دقت‌های پیش‌بینی به‌دست‌آمده در این مقاله، نتایج حاصل از داده‌های آزمایش است. در ادامه نتایج به‌دست‌آمده در هر دو الگوریتم آورده شده است.

۵-۲-۱- الگوریتم باهم‌آیی

همان‌طور که در تعاریف بخش ۳-۲ گفته شد، به‌منظور ارزیابی میزان دقت قوانین ایجادشده در الگوریتم قوانین باهم‌آیی دو معیار پشتیبان و اعتماد مورد بررسی قرار گرفته می‌شود. به‌منظور استخراج قوانین انجمنی، نیاز است حداقل مقدار پشتیبانی توسط کاربر مشخص شود. در این مقاله، یک مجموعه از حداقل پشتیبان‌ها به منظور تجزیه و تحلیل تعداد الگوهای ترتیبی استخراج‌شده در نظر گرفته شده است. برای مثال اگر حداقل مقدار پشتیبان، ۱۰ درصد در نظر گرفته شود، الگوهای حرکتی با حداقل پشتیبانی بالاتر از ۱۰ درصد استخراج می‌شود. مجموعه مقادیر پشتیبان در نظر گرفته‌شده برای استخراج الگوها به ترتیب ۳۰ درصد، ۴۰ درصد، ۵۰ درصد، ۶۰ درصد، ۷۰ درصد است. که نتایج تعداد الگوهای ترتیبی استخراج‌شده در نمودار ۳ نشان داده شده است. همان‌طور که در نمودار دیده می‌شود، با افزایش مقدار حداقل پشتیبانی تعداد الگوهای ترتیبی حرکتی کم می‌شود. علت این است که مدل الگوی قوانین باهم‌آیی نسبت به استخراج الگوها حساس‌تر شده است. اگر حداقل مقدار پشتیبانی بر این اساس قرار داده شود که الگوهای حرکتی بسیار مکرر پیدا شود، این امر سبب افزایش دقت پیش‌بینی می‌شود. اما از طرف دیگر الگوهای دیگری که می‌توانند به میزان قابل توجهی در پیش‌بینی مؤثر باشند، نادیده گرفته می‌شوند.

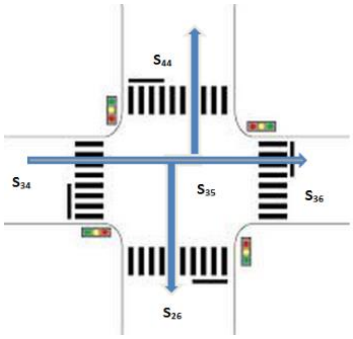


نمودار ۳: تعداد الگوهای استخراج‌شده متناسب با مقادیر پشتیبان

از دیگر معیارهای مهم در استخراج قوانین باهم‌آیی در نظر گرفتن مقدار اعتماد به‌دست‌آمده هنگام ایجاد قانون‌هاست. بر طبق تعریف‌های قبل این قانون‌ها به‌صورت مجموعه رخدادهای متوالی است که براساس رخدادهای متوالی قبلی، به‌دست می‌آیند. بعد از استخراج قوانین باهم‌آیی، نیاز است قانون‌های حرکتی برای همه زیرالگوهای مختلف که می‌توانند رخ دهند، تولید شود. توسط محاسبه مقدار اعتماد، میزان احتمال

[1] Partition Node

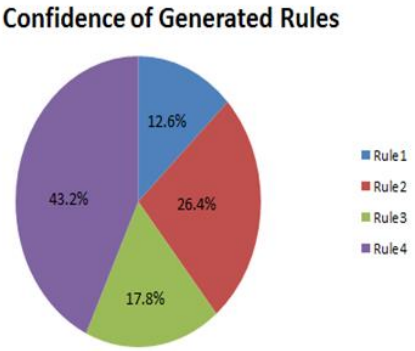
اتفاق افتادن یک رخداد معین براساس رخدادهای متوالی گذشته به دست می‌آید. برای درک بهتر اعتماد قوانین ایجادشده توسط مدل الگوی ترتیبی، تقاطع S_{35} در شکل ۷ را در نظر می‌گیریم. چهار قانون ایجادشده برای این تقاطع به صورت زیر می‌تواند باشد:



- Rule 1: $S_{34} \rightarrow S_{35} \rightarrow S_{35}$ (ماندن در تقاطع)
- Rule 2: $S_{34} \rightarrow S_{35} \rightarrow S_{44}$ (گردش به چپ)
- Rule 3: $S_{34} \rightarrow S_{35} \rightarrow S_{26}$ (گردش به راست)
- Rule 4: $S_{34} \rightarrow S_{35} \rightarrow S_{36}$ (گردش مستقیم)

شکل۷: قانون‌های حرکتی برای قطعه S35

بر طبق نتایج به‌دست‌آمده در نمودار ۴ که نشان‌دهنده مقدار اعتماد برای قانون‌هاست، خودرو به احتمال ۱۲/۶ درصد در تقاطع یعنی قطعه S_{35} می‌ماند. هم‌چنین به‌احتمال ۲۶/۴ درصد به سمت چپ گردش می‌کند، یا به احتمال ۱۷/۸ درصد به راست گردش می‌کند و به احتمال ۴۳/۲ درصد به مسیر مستقیم خود ادامه می‌دهد.



نمودار ۴: مقدار اعتماد به‌دست‌آمده برای قوانین حرکتی قطعه S35

۵-۲-۲- الگوریتم GRI

نتایج الگوریتم GRI به صورت دو عبارت مقدم و تالی است. شکل ۹ تعدادی از قوانین استخراج‌شده توسط الگوریتم را نشان می‌دهد. برای به دست آوردن این قوانین شناسه خودرو، روز حرکت خودرو و قطعه‌های قبلی طی‌شده به‌عنوان ورودی به الگوریتم داده شده است و مقصد خودرو توسط الگوریتم پیش‌بینی شده است.

Antecedent	Consequent	Support %	Confidence %
2741	2741	100.0	100.0
3047	15408	100.0	100.0
22347	4479	100.0	100.0
19230	19230	100.0	100.0
3047	4479	100.0	100.0
18556	15408	100.0	100.0
22347	15408	100.0	100.0
18556	2296	100.0	100.0
2741	15408	100.0	50.0
4478	4954	100.0	50.0
19230	15408	100.0	50.0
4954	15408	100.0	50.0
19230	4954	100.0	50.0
4479	15408	100.0	50.0
2296	15408	100.0	50.0
19230	2741	100.0	50.0
4478	15408	100.0	50.0
19230	4479	100.0	50.0
22347	3047	100.0	50.0
22347	18556	100.0	50.0
4479	2296	100.0	50.0
19230	4478	100.0	50.0
4479	4955	100.0	50.0
4478	2296	100.0	50.0
19230	2296	100.0	50.0
4478	4955	100.0	50.0
4479	4954	100.0	50.0

شکل۸ : تعدادی از قوانین به‌دست‌آمده توسط الگوریتم باهم‌آیی

نتایج به‌دست‌آمده از میزان دقت پیش‌بینی الگوریتم GRI در دو حالت قطعه بعدی و مقصد نهایی در هر دو پایگاه‌داده حالت کلی و پیچیده در نمودار۵ نشان داده شده است. نمودار۵ (الف) نشان‌دهنده میزان دقت پیش‌بینی قطعه بعدی و نمودار۵ (ب) نشان‌دهنده دقت پیش‌بینی مقصد نهایی در دو پایگاه‌داده حالت کلی و پیچیده است.

همان‌طور که در نمودار ۵ نشان داده شده است، در هر دو پایگاه‌داده حالت کلی و پیچیده با افزایش تعداد قطعه‌های طی‌شده توسط خودرو دقت پیش‌بینی

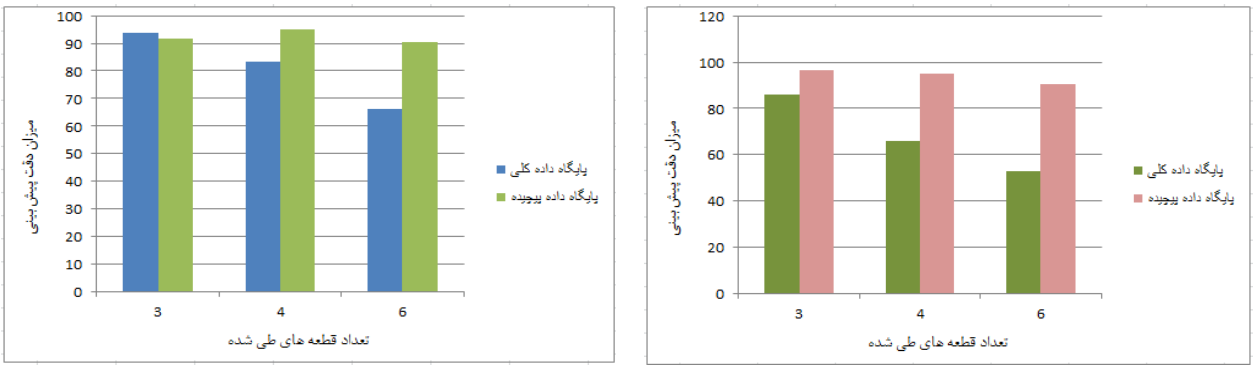
کاهش یافته است. هم‌چنین میزان دقت پیش‌بینی در پایگاه‌داده حالت پیچیده‌تر نسبت به پایگاه‌داده حالت کلی بهتر است. به همین منظور این الگوریتم برای پیش‌بینی مقصد نهایی در حالت پیچیده‌تر که تعداد قطعه کمتری در چهارراه‌ها و تقاطع در نظر گرفته می‌شود، کارایی بهتری دارد.

۵-۲-۳- الگوریتم شبکه‌های عصبی مصنوعی

همان‌طور که در بخش۳-۱ آورده شده، شبکه‌های عصبی مصنوعی در مدل‌سازی، پیش‌بینی، رده‌بندی و... با حجم داده بسیار زیاد در داده‌کاوی استفاده می‌شود. این شبکه‌ها عموماً عملکرد خود را طی یک فرآیند یادگیری فرا می‌گیرند.

```
Rule 3 for  188
    if      DATE = 02/03/2008
    and     S3 = 187
    then    188
Rule 4 for  188
    if      DATE = 02/03/2008
    and     S2 = 167
    then    188
Rule 5 for  188
    if      ID = 1374
    and     S3 = 187
    then    188
Rule 6 for  188
    if      ID = 1374
    and     S2 = 167
    then    188
Rule 7 for  188
    if      ID = 1374
    and     DATE = 02/03/2008
    and     S3 = 187
    then    188
Rule 8 for  188
    if      ID = 1374
    and     DATE = 02/03/2008
    and     S2 = 167
```

شکل ۹: تعدادی از قوانین استخراج‌شده از الگوریتم GRI



(الف)

(ب)

نمودار ۵: مقایسه میزان دقت پیش‌بینی مقصد نهایی و قطعه بعدی الگوریتم GRI در دو پایگاه‌داده حالت کلی و پیچیده‌تر

نتایج به‌دست‌آمده در جدول ۴ نشان‌دهنده میزان دقت روش‌های مختلف شبکه‌های عصبی مصنوعی به منظور انجام عملیات پیش‌بینی است، که لایه‌های داخلی نشان‌دهنده ورودی‌های شبکه و لایه‌های خارجی، آخرین قطعه طی‌شده توسط خودرو که مقصد نهایی خودرو در نظر گرفته شده است. این نتایج دقت پیش-بینی برای ۷۲۰۰ رکورد مسیر طی‌شده توسط خودروهاست، که بالاترین دقت مربوط به روش چندگانه[1] است، که میزان دقت نشان می‌دهد که این روش، ۹۶/۸۹ درصد مقصد نهایی خودروها را متناسب با مسیرهای قبلی طی‌شده به‌درستی تشخیص داده است. هم‌چنین تعداد لایه‌های پنهان متناسب با هر متد متفاوت است.

جدول۴: نتایج به‌دست‌آمده از متدهای مختلف شبکه عصبی‌مصنوعی

روش	دقت	لایه‌های داخلی	لایه‌های پنهان۱	لایه‌های پنهان۲	لایه‌های خارجی	مدت زمان اجرا [s]
پویا[2]	۸۶/۲۴	۲۲	۲۸ نرون	-	۱	۱۲۴۸
چندگانه	۹۶/۸۹	۲۲	۷۲ نرون	۱۴ نرون	۱	۴۵۴۰
هرس[3]	۹۴/۱۵	۲۲	۲۵ نرون	-	۱	۲۷۸۵

[1] Multiple

[2] Dynamic

[3] Prune

همچنین به منظور بررسی دقت پیش‌بینی مقصد نهایی متناسب با تعداد قطعه‌های متوالی طی‌شده توسط خودرو، مدل پیشنهادی شبکه‌های عصبی مصنوعی برای تعداد قطعه‌های مختلف طی‌شده در دو پایگاه‌داده حالت کلی و پیچیده، اجرا شده است. جداول ۵ و ۶ میزان دقت پیش‌بینی و زمان اجرای شبکه عصبی مصنوعی در حالت قطعه بعدی در دو پایگاه‌داده حالت کلی و پیچیده را نشان می‌دهند.

جدول ۵: دقت پیش‌بینی قطعه بعدی الگوریتم ANN در پایگاه‌داده حالت کلی

تعداد قطعه‌ها	۱۵	۱۸	۲۲
دقت پیش‌بینی	۹۵/۸۴۵	۹۶/۸۹	۹۶/۶۹
زمان اجرا [ثانیه]	۷۸۹	۴۴۵۶	۴۵۴۰

جدول ۶: دقت پیش‌بینی قطعه بعدی الگوریتم ANN در پایگاه‌داده پیچیده‌تر

تعداد قطعه‌ها	۳	۴	۶
دقت پیش‌بینی	۷۷/۰۷	۸۶/۶۴	۸۸/۷۸
زمان اجرا [ثانیه]	۱۸۵	۵۵۵	۱۵۰۰

همچنین جداول ۷ و ۸ نشان‌دهنده میزان دقت و زمان اجرای الگوریتم شبکه عصبی مصنوعی در حالت پیش‌بینی مقصد نهایی متناسب با تعداد قطعه‌های متفاوت طی‌شده توسط خودرو، در هر دو پایگاه‌داده حالت و پیچیده‌تر است.

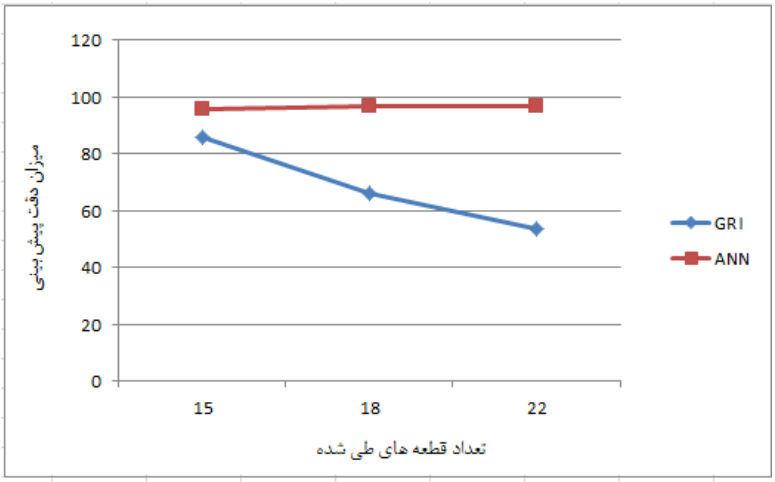
جدول ۷: دقت پیش‌بینی مقصد نهایی در الگوریتم ANN در پایگاه‌داده حالت کلی

تعداد قطعه‌ها	۷	۹	۱۲
دقت پیش‌بینی	۹۳/۸۴۵	۹۵/۸۹	۹۶/۸۹
زمان اجرا [ثانیه]	۷۸۹	۴۴۵۶	۴۵۴۰

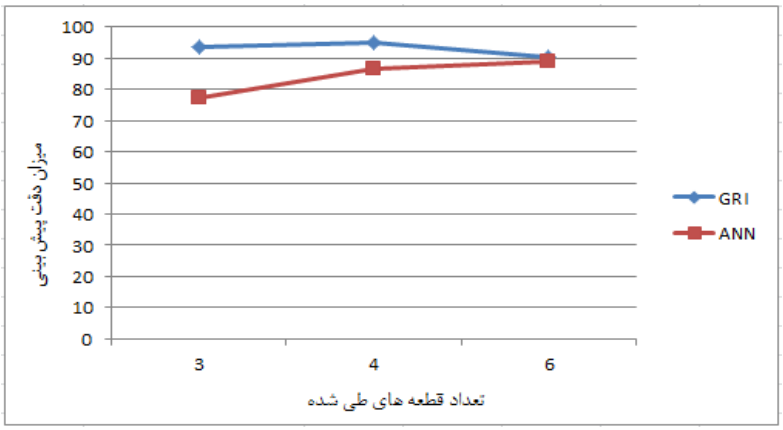
جدول ۸: دقت پیش‌بینی مقصد نهایی در الگوریتم ANN در پایگاه‌داده حالت پیچیده‌تر

تعداد قطعه‌ها	۳	۴	۶
دقت پیش‌بینی	۸۱/۶۰	۸۶/۶۴	۹۱/۶۰
زمان اجرا [ثانیه]	۲۳۴	۲۴۵	۸۶۷

مقایسه بین میزان دقت پیش‌بینی قطعه بعدی در دو الگوریتم GRI و ANN در حالت پایگاه‌داده کلی در نمودار۶ و در حالت پیچیده در نمودار۷ نشان‌داده شده‌است.

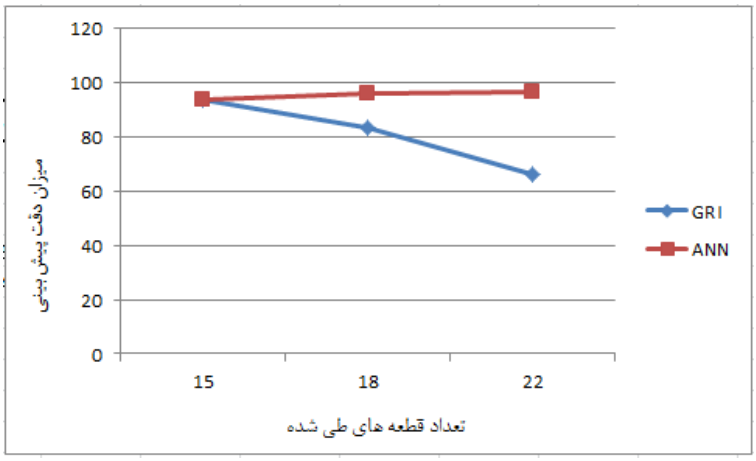


نمودار ۶: مقایسه بین دقت پیش‌بینی قطعه بعدی در الگوریتم‌های ANN و GRI در حالت کلی

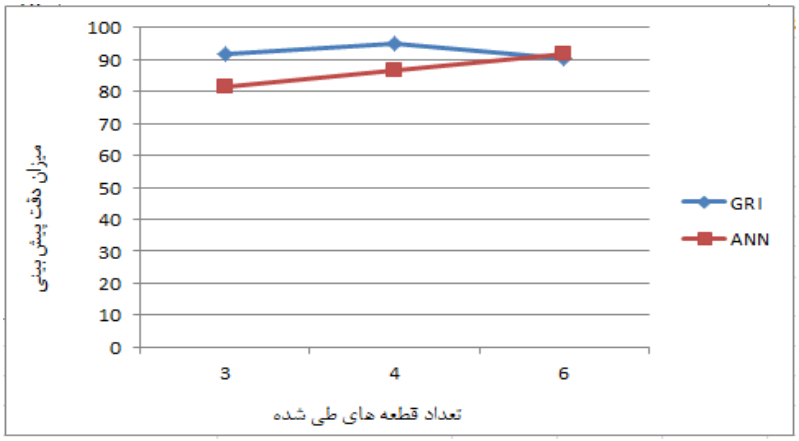


نمودار ۷: مقایسه بین دقت پیش‌بینی قطعه بعدی در الگوریتم‌های ANN و GRI در حالت پیچیده

مقایسه بین میزان دقت پیش‌بینی مقصد نهایی در دو الگوریتم GRI و ANN در حالت پایگاه‌داده کلی در نمودار ۸ و حالت پیچیده در نمودار ۹ نشان داده شده است.



نمودار ۸: مقایسه بین دقت پیش‌بینی مقصد نهایی در الگوریتم‌های ANN و GRI در حالت کلی



نمودار ۹: مقایسه بین دقت پیش‌بینی مقصد نهایی در الگوریتم‌های ANN و GRI در حالت پیچیده

همان‌طور که در نمودارهای بالا نشان داده شده است، شبکه‌های عصبی مصنوعی در هر دو پایگاه‌داده حالت کلی و پیچیده نتایج خوبی به‌دست آورده است. اما در پایگاه‌داده حالت کلی نتایج به‌دست‌آمده توسط شبکه‌های عصبی مصنوعی بهتر است و این الگوریتم با افزایش تعداد قطعه‌های پیموده‌شده توسط خودرو، دقت بیشتری به‌دست می‌آورد. به عبارت دیگر یادگیری این الگوریتم افزایش می‌یابد.

الگوریتم GRI در پایگاه داده پیچیده‌تر بهتر از شبکه‌های عصبی مصنوعی عمل کرده است، هم‌چنین زمان اجرای الگوریتم GRI بسیار کمتر از زمان اجرای مدل شبکه‌های عصبی مصنوعی است.

۶- نتیجه‌گیری و پیشنهادات

یکی از مشکلات دنیای امروز افزایش نرخ ترافیک در شهرهاست که به سبب بروز مشکلات فراوان اجتماعی، اقتصادی و آلودگی هوا شده است. تاکنون روش‌های گوناگونی به منظور کنترل ترافیک انجام شده است. در این تحقیق روش پیش‌بینی مسیر به منظور کنترل و برنامه‌ریزی ترافیکی ارائه شده است. برای رسیدن به این هدف از داده‌های واقعی GPS به‌دست‌آمده از خودروها به منظور انجام عملیات پیش‌بینی مسیر و مقصد نهایی استفاده شده است. هم‌چنین به کمک نرم‌افزار قدرتمند ArcGIS یک پایگاه‌داده دقیق از مسیرهای طی‌شده توسط خودرو استخراج شده است. به منظور انجام عملیات پیش‌بینی از الگوریتم‌های شبکه‌های عصبی مصنوعی و قوانین انجمنی استفاده شده‌است. در این مقاله از دو پایگاه‌داده حالت کلی جاده‌ها و پایگاه‌داده پیچیده‌تر که شامل چهارراه‌ها و تقاطع‌هاست به منظور ارزیابی نتایج استفاده شده است. نتایج به‌دست‌آمده نشان‌دهنده این موضوع است که شبکه‌های عصبی مصنوعی در پایگاه‌داده حالت کلی بسیار بهتر از الگوریتم GRI عمل کرده است و این مدل با افزایش تعداد قطعه‌های طی‌شده توسط خودرو نتایج بهتری به‌دست می‌آورد. اما در پایگاه‌داده حالت پیچیده‌تر تعداد قطعه‌های کمتری نسبت به حالت کلی طی می‌شود. به‌همین دلیل الگوریتم GRI بهتر از الگوریتم شبکه‌های عصبی مصنوعی عمل می‌کند. از پیشنهادات آینده می‌توان بر این موضوع تأکید کرد که با توجه به زمان اجرای بسیار پایین الگوریتم GRI نسبت به زمان اجرای الگوریتم شبکه‌های عصبی مصنوعی، بر بهبود دقت الگوریتم GRI در حالت کلی کار شود.

مراجع

[1] P. S. Castro, D. Zhang, and S. Li, "Urban traffic modelling and prediction using large scale taxi GPS traces," in Pervasive Computing, ed: Springer, 2012, pp. 57-72.

[2] S. Kamran and O. Haas, "A multilevel traffic incidents detection approach: Identifying traffic
Patterns and vehicle behaviours using real-time gps data," in Intelligent Vehicles Symposium, 2007, 2007, pp. 912-917.

[3] Y. Lou, C. Zhang, Y. Zheng, X. Xie, W. Wang, and Y. Huang, "Map-matching for low-sampling-rate GPS trajectories ",in Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, 2009, pp. 352-361.

[4] H.-w. Chang, Y.-c. Tai, and J. Y.-j. Hsu, "Context-aware taxi demand hotspots prediction," International Journal of Business Intelligence and Data Mining, vol. 5, pp. 3-18, 2009.

[5] L. Liu, C. Andris, and C. Ratti, "Uncovering cabdrivers' behavior patterns from their digital traces," Computers, Environment and Urban Systems, vol. 34, pp. 541-548, 2010.

[6] W. Liu, Y. Zheng, S. Chawla, J. Yuan, and X. Xing, "Discovering spatio-temporal causal interactions in traffic data streams," in Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining, 2011, pp. 1010-1018.

[7] D. Srinivasan ,M. C. Choy, and R. L. Cheu, "Neural networks for real-time traffic signal control," Intelligent Transportation Systems, IEEE Transactions on, vol. 7, pp. 261-272, 2006.

[8] M. Giannopoulou, Y. Roukounis, and V. Stefanis, "Traffic network and the urban environment: an adapted space syntax approach," Procedia-Social and Behavioral Sciences, vol. 48, pp. 1887-1896, 2012.

[9] M. Khalesian, P. Pahlavani, and M. R. Delavar, "A GIS-based Traffic Control Strategy Planning at Urban Intersections," IJCSNS, vol. 9, p. 166-174, 2009.

[*10*] B. Jiang and C. Liu, "Street-based topological representations and analyses for predicting traffic flow in GIS," International Journal of Geographical Information Science, vol. 23, pp. 1119-1137, 2009.

[11] Y. Li and M. McDonald, "Motorway incident detection using probe vehicles," Proceedings of the ICE-Transport, vol. 158, pp. 11-15, 2005.

[12] A. G. Evsukoff and N. F. Ebecken, "Mining fuzzy rules for a traffic information system," in Knowledge-Based Intelligent Information and Engineering Systems, 2003, pp. 237-243.

[13] K. Ozbay and P .Kachroo, "Incident management in intelligent transportation systems," 1999.

[14] J. Niittymaki and S. Kikuchi, "Application of fuzzy logic to the control of a pedestrian crossing signal," Transportation Research Record: Journal of the Transportation Research Board, pp. 30-38, 1998.

[15] J. Niittymaki and M. Maenpaa, "The role of fuzzy logic public transport priority in traffic signal control," Traffic Engineering+ Control, vol. 42, pp. 22-4, 2001.

[16] M.-S. Chen, J. Han, and P. S. Yu, "Data mining: an overview from a database perspective,", IEEE Transactions on Knowledge and data Engineering, vol. 8, pp. 866-883, 1996.

[17] M. J. Berry and G. Linoff, Data mining techniques: for marketing, sales, and customer support: John Wiley & Sons, 1997.

[18] R. Xu and D. Wunsch, "Survey of clustering algorithms," Neural Networks, IEEE Transactions on, vol. 16, pp. 645-678, 2005.

[19] W. Jian and L. X. Ming, "A Novel Algorithm for Dynamic Mining of Association Rules," in WKDD 2008. First International Workshop on Knowledge Discovery and Data Mining, 2008, pp. 94-99.

[20] K. Thearling, "An introduction to data mining," Direct Marketing Magazine, pp. 28-31, 1999.

[21] K. Swarnalatha, G. Srinivasan, and K. Kiran, "A Roadmap on Design Models for Dynamic Traffic Prediction using Multi-layer Perceptron networks".

[22] A. F. Merah, S. Samarah, and A. Boukerche, "Vehicular movement patterns: a prediction-based route discovery technique for VANETs," in *Communications (ICC), 2012 IEEE International Conference on*, 2012, pp. 5291-5295.

[23] R. Agrawal and R. Srikant, "Mining sequential patterns," in *Data Engineering, 1995. Proceedings of the Eleventh International Conference on*, 1995, pp. 3-14.

[24] M. W. Sermons and F. S. Koppelman, "Use of vehicle positioning data for arterial incident detection," Transportation Research Part C: Emerging Technologies, vol. 4, pp. 87-96, 1996.

[25] C. C. Aggarwal, Data streams: models and algorithms vol. 31: Springer Science & Business Media, 2007.

[26]http://research.microsoft.com/apps/pubs/?id=152883

[27]http://download.bbbike.org/osm/bbbike/Beijing/

K. N. TOOSI UNIVERSITY OF TECHNOLOGY
FACULTY OF GEODESY AND GEOMATICS ENGINEERING

Journal of Geospatial Information Technology
Vol.3, No.2, Summer 2015

Forecasting Traffic Load using GPS data, a data mining approach

Zahra Mahdavian*[1], Aliakbar Niknafs[2]

1--Master of Information Technology Engineering, Department of Electronic and Computer Engineering; Graduate University of Advanced Technology, Kerman, Iran
2-Assistant Professor; Faculty of Computer Engineering Department; Shahid Bahonar University of Kerman, Kerman, Iran

Abstract

In today's world, rapid increase of urbanization and traffic challenges have led to the profound need of traffic control systems with the highest possible productivity and efficiency. Time loss and increased fuel consumption as well as air and noise pollutions have made traffic control to be one of the most important current issues in the world. One of the best possible methods for reaching this objective is to predict the directions and the final destination of the car. If the future position of a car can be predicted, traffic estimation in an urban zone will be a simple task. Route prediction is possible based on the previous routes of the car as well as parameters such as the starting point, time, day, month, and duration utilizing data mining methods and artificial neural networks .The current paper uses real GPS data obtained from different cars in order to carry out prediction operation of the route and final destination. One of the propoesd methods in this study is to establish a database for the previous routes of the cars using ArcGIS software The high accuracy of recording the previous routes of the cars in this database increases the accuracy of the route prediction process.In this study, two distinct databases were established. The first involves general database, where only the more challenging sections of the roads including intersections and crossroads are considered in the second which is the more complex database.Moreover, in order to carry out the prediction operation, association rules algorithms as well as artificial neural network algorithms have been used. The obtained results indicate the high accuracy of the prediction. Artificial Neutral Network (ANN) algorithms used on the general database and the GRI algorithm used on the more complex one provide better results, respectively. Both algorithms acquires precision greater than 95%. The results obtained from the prediction process can be used for traffic planning and the optimization of car movements.

Key words: Traffic, Data Mining, Prediction , GPS

Correspondence Address: Department of Electrical and Computer Engineering, Kerman Graduate University of Technology, Kerman, Iran.
Tel: +98 34 33776611-13.
Email: z.mahdavian@student.kgut.ac.ir