

استخراج ویژگی از خطوط سیر به منظور شناسایی حالت حمل و نقل در شهر

هوشمند با استفاده از یادگیری ماشین

سجاد صولتی، دانشجوی کارشناسی ارشد سیستم‌های اطلاعات مکانی، پردیس دانشکده فنی، دانشگاه تهران، تهران، ایران
رحیم علی عباسپور، دانشیار دانشکده مهندسی نقشه برداری و اطلاعات مکانی، پردیس دانشکده‌های فنی، دانشگاه تهران، تهران، ایران
علیرضا چهرقان (مسئول مکاتبات)، استادیار دانشکده مهندسی معدن، دانشگاه صنعتی سهند، تبریز، ایران

E-mail: chehrehgan@sut.ac.ir

پذیرش: ۱۴۰۰/۰۴/۱۹

دریافت: ۱۴۰۰/۰۲/۰۱

چکیده

با پیشرفت فناوری و اهمیت حمل و نقل هوشمند در شهر هوشمند، پیش‌بینی و شناسایی استفاده از حالات حمل و نقلی بخش اولیه و اساسی در حمل و نقل هوشمند نظیر مسائل کاربردی کنترل ترافیک، تجزیه و تحلیل تقاضای سفر و برنامه‌ریزی حمل و نقل به شمار می‌آید. با فراگیر شدن دستگاه‌های تعیین موقعیت و همچنین استفاده از تلفن‌های هوشمند، مسیرهای موقعیتی ثبت شده یک رویکرد اقتصادی و سریع به منظور شناسایی حالات حمل و نقل به حساب می‌آید. در این پژوهش با استفاده از داده‌های موقعیتی ثبت شده توسط مرکز تحقیقاتی میکروسافت آسیا (GeoLife) که بیشتر در شهر پکن انجام پذیرفته و استخراج ویژگی‌های حرکتی به پیش‌بینی حالات حمل و نقلی پیاده‌روی، استفاده از دوچرخه، استفاده از اتوبوس، استفاده از اتومبیل و استفاده از قطار پرداخته می‌شود. رویکرد پیشنهادی استفاده از کلاس‌بندی‌های یادگیری ماشین، شامل: کلاس‌بندی ماشین بردار پشتیبان، جنگل تصادفی، تقویت گرادیان و تقویت بیش‌ازحد گرادیان می‌باشد. در بین مدل‌ها، مدل تقویت بیش‌ازحد گرادیان توانسته با کسب دقت ۹۵/۱۸ درصد و پیچیدگی زمانی بهتر با توجه به سایر مدل‌ها، حالات حمل و نقلی را با دقت بالاتری پیش‌بینی کند.

کلمات کلیدی: پیش‌بینی حالات حمل و نقل، حمل و نقل هوشمند، مجموعه داده GeoLife، یادگیری ماشین

۱. مقدمه

ایده شهر هوشمند در اوایل سال ۱۹۹۰ میلادی بیان شد، این واژه به معنای توسعه شهرها همراه با پیشرفت فناوری و نوآوری در جهان است (Tamane, Dey et al. 2020). گسترش فناوری جنبه اصلی شهرهوشمند می‌باشد (Jarašūniene 2007)، شهرهوشمند زمانی تاثیر دارد که بتواند با بکارگیری فناوری‌های نوین موجب ارزش عمومی برای آحاد شهروندان گردد (Chamoso, González-Briones et al. 2018). زیرساخت‌های موجود در هر شهر هوشمندی می‌بایست همواره بروز باشند، شهرها با مشکلات ناکافی بودن حمل‌ونقل عمومی، ترافیک و انتشار گازهای گلخانه‌ای ناشی از حمل‌ونقل روبرو هستند (Tamane, Dey et al. 2020) و وجود این موارد ضرورت توجه به ایجاد حمل‌ونقل هوشمند در این شهرهاست. عادات حرکتی افراد مطابق با یک چرخه زمانی معین دارای نظم و الگوی خاصی است، شناخت این الگوهای پنهان یکی از مهمترین مراحل توسعه و اداره شهرها در حمل‌ونقل هوشمند می‌باشد (Adler and Blue 1998, Scheiner and Holz- 2007). با محاسبه نحوه تغییر مکان افراد در بین نقاط مورد علاقه^۱ می‌توان به آشکارسازی این الگوها کمک کرد. شناسایی حالات حمل‌ونقل^۲ یک گام اساسی برای حل مشکلات حمل‌ونقلی نظیر تجزیه و تحلیل تقاضای سفر، برنامه‌ریزی حمل‌ونقلی و مدیریت ترافیک می‌باشد (Zheng, Chen et al. 2010, Dabiri and Heaslip 2018). در گذشته اطلاعات پیش‌بینی و شناسایی نوع حمل‌ونقل افراد به صورت فیزیکی با ارائه پرسشنامه و یا تماس تلفنی ثبت می‌شد (Wu, Yang et al. 2016, Xiao, Wang et al. 2017, Dabiri and Heaslip 2018)، این اطلاعات به شکل جهانی نبوده و کیفیت داده‌ها بسیار تحت تاثیر پاسخ دهندگان قرار داشت که معمولاً نادرست و ناقص بودند (Yang, Stewart et al. 2018). پیشرفت فناوری نوبری، سیستم موقعیت‌یابی جهانی، خدمات مبتنی بر مکان، سیستم اطلاعات مکانی و گسترش تلفن-های همراه هوشمند در کنار توسعه سریع شهرها منجر به تمرکز

استفاده از سیستم‌های حمل‌ونقل در جهان امروز نقش کلیدی در زندگی روزمره انسان‌ها دارد و توسعه آن با توجه به جایگاه و اهمیت حمل‌ونقل در کشور ضروری است. افزایش جمعیت و تنوع وسایل حمل‌ونقل در چند دهه گذشته زندگی شهروندان را تحت تاثیر خود قرار داده است. توسعه شهرها و مشکلات گسترده شهرنشینی سبب شده تا شهروندان بیش از پیش نیاز به تغییر مکان سریع داشته باشند و این امر موجب ایجاد ترافیک، تصادفات، آلودگی هوا و مصرف انرژی می‌شود. ساخت و احداث راه، یکی از ابتدایی‌ترین روش‌ها برای مقابله با این موضوع است. احداث راه علاوه بر پرهزینه و زمان‌بر بودن موجب تخریب محیط‌زیست می‌گردد، همچنین اجرای آن نیازمند فضای ساخت گسترده در مناطق احداث خواهد بود که در نظر گرفتن این موارد در فضای شهری دشوار است (Jarašūniene 2006). اگرچه افزایش وسایل نقلیه در محیط شهری ممکن است جنبه منفی داشته باشد، اما در صورت برنامه‌ریزی صحیح، این وسایل می‌توانند به عنوان ابزاری برای نظارت هوشمند با استفاده از فناوری‌های جدید مورد استفاده قرار گیرند. عدم برنامه‌ریزی توسعه مراکز شهری علاوه بر موارد ذکر شده موجب مشکلات اقتصادی و اجتماعی شدیدی خواهد شد (Zhong, Sumalee et al. 2013, Brennand, Maia et al. 2019). به همین دلیل نیاز به یک سیستم هوشمند حمل‌ونقلی بیش از پیش مورد توجه کشورهای توسعه‌یافته قرار گرفته است. طبق بیانیه EU/۲۰۱۰/۴۰ اتحادیه اروپا، سیستم حمل‌ونقل هوشمند به عنوان سیستمی تعریف می‌شود که در آن از فناوری‌های اطلاعاتی و ارتباطی در زمینه حمل‌ونقل به منظور مدیریت زیرساخت‌ها، وسایل نقلیه، کاربران، جریان‌های ترافیکی و حرکتی استفاده شده باشد که در ارتباط با حالات حمل‌ونقلی دیگر نیز باشد (Payne 2015). حمل‌ونقل هوشمند یکی از اجزای مؤثر در شهر هوشمند به حساب می‌آید. اولین

استخراج ویژگی از خطوط سیر به منظور شناسایی حالت حمل و نقل در شهر هوشمند با استفاده از یادگیری ماشین

بر ایجاد روش‌های جدیدی برای شناسایی حالات حمل و نقل شده است. در سال‌های اخیر، گسترش فناوری GPS^۲ توانایی ثبت حجم عظیمی از داده‌ها با صرف هزینه کم، سرعت بالا و درصد خطای پایین در کمترین زمان را فراهم آورده است (Murakami and Wagner 1999, Dabiri and Heaslip 2018, Dabiri and Heaslip 2018, Huang, Wang et al. 2020). استفاده از این داده‌ها فرصت‌های زیادی برای پژوهشگران در حوزه‌های مختلف ایجاد کرده است. یکی از موارد استفاده از داده‌های GPS در شناسایی حالات حمل و نقل می‌باشد. این پژوهش با استفاده از داده‌های GeoLife به عنوان داده‌های GPS مورد پردازش استفاده شده (Zheng, Fu et al. 2011) به استخراج حالات حمل و نقل شامل: پیاده‌روی، استفاده از دوچرخه، استفاده از اتومبیل، استفاده از اتوبوس و استفاده از قطار را می‌پردازد. تعداد چهار ویژگی نقطه‌ای کینماتیک، چهارده ویژگی آماری سفر برای هر یک از ویژگی‌های نقطه‌ای به همراه سه ویژگی پیشرفته استخراج می‌شود. به منظور افزایش دقت، ویژگی‌های مهم سفر با اجرای روش ترکیبی انتخاب شده‌اند. پیاده‌سازی نیز با استفاده از چهار مدل کلاسه-بندی یادگیری ماشین شامل مدل‌های ماشین بردار پشتیبان^۳، جنگل تصادفی^۴، تقویت گرادیان^۵ و تقویت بیش‌ازحد گرادیان^۶ انجام شده است.

در ادامه، بخش دوم به بررسی و مرور مطالعات انجام شده در شناسایی حالات حمل و نقل می‌پردازد. بخش سوم به بیان تعاریف و رویکرد کلی مسئله اختصاص یافته و در بخش چهارم پیاده‌سازی و ارزیابی نتایج انجام می‌گیرد. بخش آخر نیز به نتیجه‌گیری و ارائه پیشنهادها می‌پردازد.

۲. مروری بر مطالعات انجام شده

مطالعات مختلفی به وسیله داده‌های ثبت شده توسط ابزارهای تعیین موقعیت، شتاب‌سنج، ژيروسکوپ و همچنین لایه‌های مکانی به منظور بالابردن دقت پیش‌بینی انجام شده است. در

(Zheng, Li et al. 2008) با استفاده از داده‌های GPS به تنهایی و اجرای مدل‌های کلاسه‌بندی جنگل تصادفی، ماشین بردار پشتیبان، CRF و BYS توانسته حالات حمل و نقلی پیاده، اتومبیل، اتوبوس و دوچرخه را با دقت ۷۲/۶ درصد پیش‌بینی کند. در (Stenneth, Wolfson et al. 2011) با در نظر گرفتن داده‌های GPS به همراه داده‌های GIS و پیاده‌سازی مدل‌های بیزساده، شبکه بیز، درخت تصمیم و جنگل تصادفی توانسته حالات حمل و نقلی شامل: پیاده، اتومبیل، اتوبوس، قطار، دوچرخه و ایستاده را با دقت ۹۳/۵ درصد پیاده‌سازی و پیش‌بینی کند. در (Jahangiri and Rakha 2015) با استفاده از داده-های شتاب‌سنج، ژيروسکوپ و بردار چرخش و اجرای کلاسه-بندی‌های ماشین بردار پشتیبان، نزدیکترین همسایگی، درخت تصمیم، جنگل تصادفی و Bag توانسته حالات پیاده، دوچرخه، اتومبیل، اتوبوس و دویدن را با دقت ۹۵/۱ درصد پیش‌بینی کند. در (Dabiri and Heaslip 2018) با استفاده از داده‌های GPS و استخراج چهار ویژگی نقطه‌ای، تشکیل یک شبکه عصبی CNN داده و توانسته با دقت ۸۴/۸ درصد حالات پیاده، دوچرخه، اتوبوس، اتومبیل و قطار را پیش‌بینی کند. در (Yu et al. 2019) به این مهم پرداخته شد که انتخاب دستی ویژگی‌ها نمی‌تواند به طور کامل الگوهای حالات حمل و نقلی را مشخص کند. لذا از الگوریتم‌های یادگیری عمیق استفاده شده که نیازمند ورودی با اندازه مشخص می‌باشد، همچنین از یک روش مبتنی بر تبدیل موجک گسسته به همراه یادگیری عمیق مبتنی بر حافظه کوتاه مدت استفاده شده است که از ویژگی‌های سرعت، شتاب، جرک^۸، چرخش و ویژگی‌های استخراج شده از موجک در حوزه فرکانس استفاده می‌کند و توانسته با استفاده از داده‌های GPS به دقت ۹۲/۷ درصد برسد. در (Nawaz, Zhiqiu et al. 2020) از الگوریتم ConvLSTM استفاده شده است تا بتوان روابط بین مکان و زمان را در نظر گرفت. در این تحقیق چهار حالت حمل و نقلی شامل پیاده، دوچرخه، اتوبوس و اتومبیل مورد بررسی قرار گرفته است. ویژگی‌های حرکتی سرعت،

بر ایجاد روش‌های جدیدی برای شناسایی حالات حمل و نقل شده است. در سال‌های اخیر، گسترش فناوری GPS^۲ توانایی ثبت حجم عظیمی از داده‌ها با صرف هزینه کم، سرعت بالا و درصد خطای پایین در کمترین زمان را فراهم آورده است (Murakami and Wagner 1999, Dabiri and Heaslip 2018, Dabiri and Heaslip 2018, Huang, Wang et al. 2020). استفاده از این داده‌ها فرصت‌های زیادی برای پژوهشگران در حوزه‌های مختلف ایجاد کرده است. یکی از موارد استفاده از داده‌های GPS در شناسایی حالات حمل و نقل می‌باشد. این پژوهش با استفاده از داده‌های GeoLife به عنوان داده‌های GPS مورد پردازش استفاده شده (Zheng, Fu et al. 2011) به استخراج حالات حمل و نقل شامل: پیاده‌روی، استفاده از دوچرخه، استفاده از اتومبیل، استفاده از اتوبوس و استفاده از قطار را می‌پردازد. تعداد چهار ویژگی نقطه‌ای کینماتیک، چهارده ویژگی آماری سفر برای هر یک از ویژگی‌های نقطه‌ای به همراه سه ویژگی پیشرفته استخراج می‌شود. به منظور افزایش دقت، ویژگی‌های مهم سفر با اجرای روش ترکیبی انتخاب شده‌اند. پیاده‌سازی نیز با استفاده از چهار مدل کلاسه-بندی یادگیری ماشین شامل مدل‌های ماشین بردار پشتیبان^۳، جنگل تصادفی^۴، تقویت گرادیان^۵ و تقویت بیش‌ازحد گرادیان^۶ انجام شده است.

در ادامه، بخش دوم به بررسی و مرور مطالعات انجام شده در شناسایی حالات حمل و نقل می‌پردازد. بخش سوم به بیان تعاریف و رویکرد کلی مسئله اختصاص یافته و در بخش چهارم پیاده‌سازی و ارزیابی نتایج انجام می‌گیرد. بخش آخر نیز به نتیجه‌گیری و ارائه پیشنهادها می‌پردازد.

۲. مروری بر مطالعات انجام شده

مطالعات مختلفی به وسیله داده‌های ثبت شده توسط ابزارهای تعیین موقعیت، شتاب‌سنج، ژيروسکوپ و همچنین لایه‌های مکانی به منظور بالابردن دقت پیش‌بینی انجام شده است. در

(2019) به منظور افزایش توانایی شناسایی نوع حمل و نقل استفاده شده است.

۳. رویکرد حل مسأله

رویکرد شناسایی حالات حمل و نقل شامل چهار گام پردازشی می‌باشد. در این روند، مطابق با فلوچارت شکل (۱) داده‌های GPS پیش‌پردازش شده و ویژگی‌های مؤثر استخراج و انتخاب می‌شوند. در ادامه مدل‌های کلاسه‌بندی پیاده‌سازی شده و در پایان مدل‌ها ارزیابی می‌شوند. در ادامه این مراحل مورد تحلیل و بررسی قرار می‌گیرند.

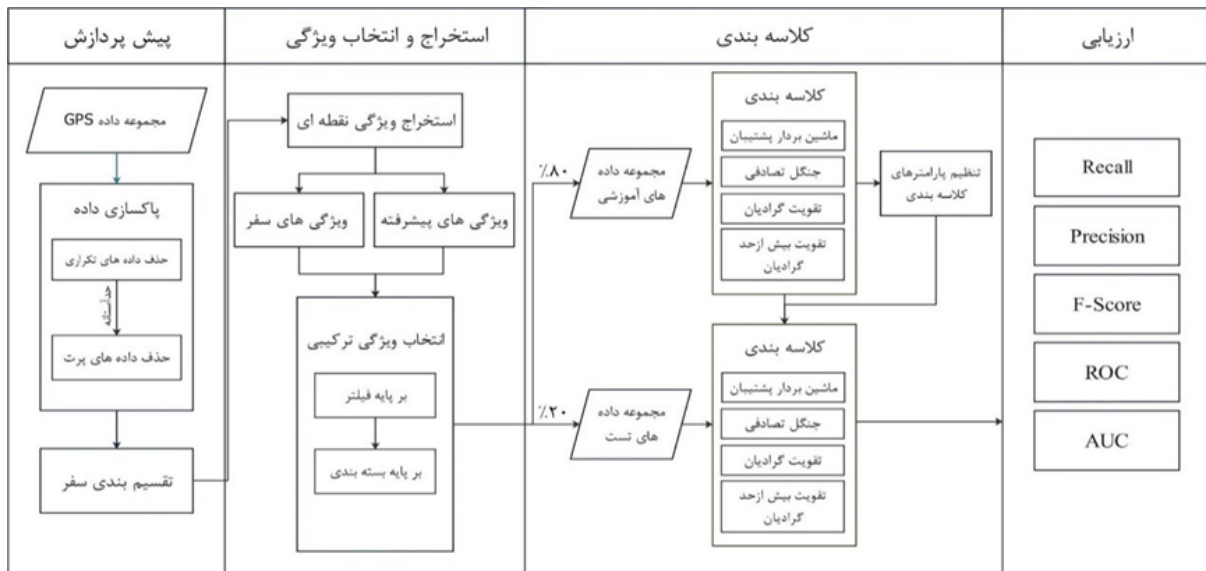
۳-۱- پیش پردازش

فرآیند پیش‌پردازش به منظور حذف داده‌های پرت و پردازش محاسبات به صورت صحیح و باکیفیت انجام می‌گیرد (Han, Kamber et al. 2011). داده‌های GPS در حوزه شناخت حالات حمل و نقل به دلیل امکان وجود اندازه‌گیری نادرست ناشی از محدودیت تکنیک‌های تعیین موقعیت، احتمال وجود خطا در دستگاه‌های ثبت نقاط و خطاهای انسانی نیاز به پیش‌پردازش و پاکسازی داده خواهد داشت (Langley 1997). یکی از روش‌های مقابله با این موضوع فیلترکردن داده‌ها بر اساس تعیین حد‌آستانه سرعت و شتاب برای هر حالت حمل و نقلی می‌باشد. در این روش باتوجه به جدول (۱) براساس ویژگی‌های طبیعی موجود در هر حالت حمل و نقل، بیشینه مقدار سرعت و شتاب برای همه حالات حمل و نقلی در نظر گرفته شده است (Dabiri and Heaslip 2018).

با توجه به رویکرد کلاسه‌بندی، می‌بایست هر سفر انجام شده فقط یک حالت حمل و نقلی را دارا باشد به همین منظور سفرهای بدست‌آمده از داده‌های پاکسازی شده به یک حالت حمل و نقلی تقسیم می‌شوند. موارد انجام شده به منظور پیش‌پردازش داده‌ها شامل مراحل زیر می‌باشد:

شتاب، جرک، تغییر جهت و دو ویژگی روزهفته و زمان برش خورده به عنوان ورودی در الگوریتم استفاده شده است. این تحقیق توانسته با اضافه کردن ویژگی‌ها، دقت را ۳ درصد بهبود بخشد. به صورت کلی شناخت نوع حمل و نقل شامل دو مرحله اساسی، استخراج ویژگی و مدل کلاسه‌بندی می‌باشد. در بیشتر تحقیقات از داده‌های GeoLife به دلیل کیفیت و کمیت بالا استفاده شده است (Zheng, Li et al. 2008, Zheng, Liu et al. 2010, Zheng, Chen et al. 2008, et al. 2010). در برخی از مطالعات نیز داده‌ها به صورت مستقل جمع‌آوری شده (Bantis and Haworth 2017) و یا توسط سازمان‌های دولتی منتشر شده‌اند (Biljecki, Ledoux et al. 2013). در (Zheng, Liu et al. 2008) از ویژگی‌های آماری متداول، شامل سرعت متوسط، سرعت انتظار، سه سرعت بالا و سه شتاب بالا استفاده شده است. در (Zheng, Li et al. 2008) سه ویژگی پیشرفته، نرخ تغییر جهت، نرخ توقف و نرخ تغییر سرعت استفاده شده که منجر به بهبود دقت مدل شده است. در (Dodge, Weibel et al. 2009) از پنجاه‌هشت ویژگی جهانی و محلی استفاده شده است. همچنین با ترکیب ویژگی‌ها و انتخاب مدل‌های درخت تصمیم (Zheng, Liu et al. 2008, Reddy, Mun et al. 2010)، نزدیکترین همسایه (Jahangiri and Rakha 2015, Zheng, Liu et al. 2008)، ماشین بردار پشتیبان (Bolbol, Cheng et al. 2012, Jahangiri and Rakha 2015, Jahangiri and Rakha 2014)، سیستم‌های فازی (Biljecki, Ledoux et al. 2013, Das and Winter et al. 2016)، کلاسه‌بندی یادگیری گروه (Jahangiri and Rakha 2014, Jahangiri and Rakha 2015, Zhu, Zhu et al. 2016, Xiao, Wang et al. 2017) و روش‌های یادگیری عمیق (Endo, Toda et al. 2016, Song, Kanasugi et al. 2016, Vu, Dung et al. 2016, Dabiri and Heaslip 2018, Guo, Wang et al.

استخراج ویژگی از خطوط سیر به منظور شناسایی حالت حمل و نقل در شهر هوشمند با استفاده از یادگیری ماشین

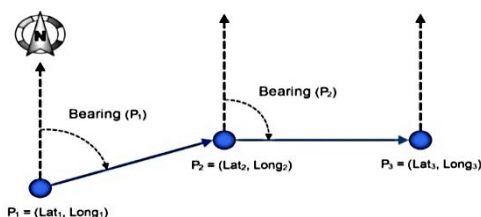


شکل ۱. فلوجارت رویکرد پیشنهادی برای حل مسأله

شناسایی حالات حمل و نقل تنها با استفاده از مختصات نقاط GPS امکان پذیر نیست و نیازمند استخراج ویژگی از نقاط ثبت شده می باشد. استخراج ویژگی های مناسب ارتباط مستقیمی با تشخیص نوع حمل و نقل دارد، لذا در این پژوهش سعی شده استخراج ویژگی های زیادی انجام شود. ویژگی ها در این مطالعه به ویژگی های نقطه ای، سفر و پیشرفته تقسیم بندی شده اند. نحوه استخراج ویژگی ها در ادامه به تفصیل بیان شده اند.

۳-۲-۱ ویژگی های نقطه ای

ویژگی های نقطه ای با توجه به مختصات نقاط در سری نقاط ثبت شده محاسبه می گردد. فاصله محاسبه شده بین نقاط بر اساس بیضوی ژئودتیک جهانی ۱۹۸۴ (WGS84) می باشد (Vincnty 1975). شکل (۲) نشان دهنده خط سیر مسیر نقاط GPS می باشد، که با استفاده از آن ویژگی های نقطه ای استخراج شده اند. هر یک از ویژگی های نقطه ای و نحوه محاسبه آن در ادامه بررسی می شود.



جدول ۱. حد آستانه سرعت و شتاب بر اساس حالات حمل و نقل

حالت حمل و نقل	بیشینه سرعت (m/s)	بیشینه شتاب (m/s ²)
پیاده	۷	۳
دوچرخه	۱۲	۳
اتوبوس	۳۴	۲
اتومبیل	۵۰	۱۰
قطار	۳۴	۳

- داده هایی که تغییرات زمانی آن با نقطه بعدی بیش از ۱۵ دقیقه باشد در یک سفر جداگانه بررسی می شوند (Guo, Liang et al. 2020).
- داده هایی که سرعت و شتاب آن با توجه به نوع حمل و نقل، بیش از حد آستانه تعیین شده باشد، حذف می شوند (Dabiri and Heaslip 2018).
- به دلیل محاسبه جرک سفرهایی که کمتر از ۳ نقطه داده دارند، حذف می شوند.
- تقسیم بندی سفرها به گونه ای که هر سفر دارای یک حالت حمل و نقل باشد.

۳-۲ استخراج ویژگی

شکل ۲. خط سیر مسیر نقاط GPS

مقادیر متفاوتی را دربر می‌گیرد، به عنوان مثال، اتومبیل و اتوبوس در طول یک خط حرکت می‌کنند و تغییر جهت کمتری دارند، در حالی که استفاده از دوچرخه و حالت پیاده، تغییر جهت بیشتری را داراست (Zheng, Li et al. 2008). محاسبه این مقدار ملزم به محاسبه تغییرات جهت در دو جهت X و Y می‌باشد، تغییر جهت بر اساس معادله (۴) محاسبه می‌شود (Dabiri and Heaslip 2018).

$$\begin{aligned} \text{Bearing Rate}_{(P_1)} &= \left| \text{Bearing}_{P_2} - \text{Bearing}_{P_1} \right| \\ \text{Bearing}_{(P_i)} &= \text{Arctan}\left(\frac{y_i}{x_i}\right) \end{aligned} \quad (4)$$

در معادله (۴) Bearing Rate_(P₁) بیانگر تغییر جهت در نقطه P₁ و Bearing_(P₁) جهت نقطه P₁ را نشان می‌دهد.

۲-۲-۳ ویژگی‌های سفر

ویژگی‌های سفر از جمله داده‌های آماری است که با توجه به نقاط GPS ثبت شده در کل سفر توانایی استخراج ویژگی‌های مؤثر را دارا است (Xiao, Wang et al. 2017). ویژگی‌های سفر محاسبه شده در این تحقیق در ادامه بیان می‌شود.

- میانگین: معیار گرایش داده‌ها به مرکز می‌باشد.
- مد: بیانگر مقداری است که بیشترین تکرار را در داده‌ها دارد.
- انحراف معیار: این مقدار بیانگر پراکندگی داده‌ها در طول سفر می‌باشد.
- بیشینه مقدار: بیشترین مقداری که در داده‌ها یافت می‌شود.
- کمینه مقدار: کمترین مقداری که در داده‌ها یافت می‌شود.
- محدوده مقدار: تفاضل بیشینه و کمینه مقدار است.
- صدک: مقداری که درصد خاصی از نمونه‌های یک متغیر تصادفی کمتر از آن مقدار است. در این مقاله، مقادیر ۱۰ فصلنامه مهندسی حمل و نقل / سال چهاردهم / شماره اول (۵۴) / پاییز ۱۴۰۱

• سرعت

ویژگی سرعت از جمله ویژگی‌های نقطه‌ای می‌باشد که در شناسایی نوع حمل و نقل کاربرد دارد. معادله (۱) نحوه محاسبه ویژگی سرعت را نشان می‌دهد (Dabiri and Heaslip 2018).

$$V_{p_1} = \frac{\text{Distance}(P_1, P_2)}{\Delta T} \quad (1)$$

در معادله (۱) V_{p_1} بیانگر سرعت در نقطه P₁، Distance(P₁, P₂) معادل فاصله ژئودتیک بین دو نقطه P₁ و P₂ و ΔT زمان طی شده بین دو نقطه بر حسب ثانیه می‌باشد.

• شتاب

دیگر ویژگی نقطه‌ای استخراج شده ویژگی شتاب می‌باشد. در این ویژگی، شتاب با توجه به محاسبه تفاضل سرعت متحرک هر زوج نقطه نسبت به زمان طی شده محاسبه می‌گردد. معادله (۲) نحوه محاسبه ویژگی شتاب را نشان می‌دهد (Dabiri and Heaslip 2018).

$$A_{p_1} = \frac{V_{p_2} - V_{p_1}}{\Delta T} \quad (2)$$

در معادله (۲) A_{p_1} بیانگر شتاب در نقطه P₁، V_{p_1} سرعت در نقطه P₁، V_{p_2} سرعت در نقطه P₂ و ΔT زمان طی شده بین دو نقطه می‌باشد.

• جرک

این ویژگی نقطه‌ای بیانگر مشتق دوم سرعت می‌باشد، این مقدار عامل مهمی در تعادل و احساس آرامش مسافران در وسایل نقلیه عمومی می‌باشد (Bagdadi and Várhelyi 2013). معادله (۳) نحوه محاسبه ویژگی جرک را نشان می‌دهد (Dabiri and Heaslip 2018).

$$J_{p_1} = \frac{A_{p_2} - A_{p_1}}{\Delta T} \quad (3)$$

که J_{p_1} بیانگر جرک در نقطه P₁، V_{p_1} سرعت در نقطه P₁، V_{p_2} سرعت در نقطه P₂ و ΔT زمان طی شده بین دو نقطه است.

- تغییر جهت: این ویژگی با توجه به حالات حمل و نقلی

استخراج ویژگی از خطوط سیر به منظور شناسایی حالت حمل و نقل در شهر هوشمند با استفاده از یادگیری ماشین

که حالت پیاده روی و دوچرخه سواری نیاز به تغییر جهت و همچنین پیچیدگی در مسیر دارد. معادله (۸) نحوه محاسبه نرخ تغییر جهت را نشان می دهد (Zheng, Li et al. 2008).

$$HCR = \frac{|P_c|}{Distance} \quad (8)$$

در معادله (۸)، $|P_c|$ بیانگر تعداد دفعاتی است که مقدار جهت نقاط GPS از حد آستانه تعیین شده بیشتر باشد. HCR را می توان تعداد تکرار تغییر جهت حرکت در فاصله طی شده دانست.

- نرخ توقف: نرخ توقف بیانگر تعداد نقاط GPS است که سرعت آن کمتر از حد آستانه تعیین شده در مسافت طی شده می باشد. در یک مسافت مشابه افرادی که از اتوبوس استفاده می کنند تعداد دفعات بیشتری نسبت به رانندگی با اتومبیل متوقف می شوند. معادله (۹) نحوه محاسبه نرخ توقف را نشان می دهد (Zheng, Li et al. 2008).

$$SR = \frac{|P_s|}{Distance} \quad (9)$$

در معادله (۹)، $P_s = \{p_i | p_i \in P, V(p_i) < V_s\}$ و V_s حد آستانه تعیین شده این ویژگی می باشد. به طور کلی نرخ توقف حالت پیاده روی بیشتر از استفاده از اتوبوس و نرخ توقف استفاده از اتوبوس بیشتر از استفاده از اتومبیل می باشد.

- نرخ تغییر سرعت: محاسبه نرخ تغییر سرعت نیز به دلیل امکان وجود شرایط خاص ترافیکی و آب و هوایی می باشد. برای محاسبه نرخ تغییر سرعت ابتدا نرخ سرعت محاسبه شده و سپس نرخ تغییرات سرعت مطابق با معادله (۱۰) محاسبه می گردد (Zheng, Li et al. 2008).

$$VCR = \frac{|P_v|}{Distance}, \quad VR(p_1) = \frac{|V_{p_2} - V_{p_1}|}{V_{p_1}} \quad (10)$$

در معادله (۱۰)، $P_v = \{p_i | p_i \in P, VR(p_i) > V_r\}$ و V_r حد آستانه تعیین شده این ویژگی می باشد. به طور کلی این مقدار بیانگر تعداد نقاط GPS است که تغییر سرعت آن بیشتر از حد آستانه تعیین شده می باشد.

امین صدک، ۲۵ امین صدک، ۷۵ امین صدک و ۹۰ امین صدک محاسبه شده اند.

- محدوده بین چارکی: تفاضل بین صدک بالا و صدک پایین را محاسبه می کند.

- چولگی: این ویژگی میزان عدم تقارن توزیع احتمال را مطابق با معادله (۵) می دهد.

$$S = E \left[\left(\frac{X - \mu}{\sigma} \right)^3 \right] \quad (5)$$

در معادله (۵)، μ بیانگر میانگین و σ انحراف معیار داده ها می باشد.

- کشیدگی: این ویژگی میزان قله ای و یا مسطح بودن توزیع احتمال را نشان می دهد. معادله (۶) نحوه محاسبه کشیدگی را نشان می دهد.

$$K = E \left[\left(\frac{X - \mu}{\sigma} \right)^4 \right] - 3 \quad (6)$$

در معادله (۶)، μ بیانگر میانگین و σ انحراف معیار می باشد.

- ضریب تغییرات: این ویژگی بدون بعد، نمایانگر میزان پراکندگی به ازای یک واحد از میانگین می باشد. نحوه محاسبه ضریب تغییرات در معادله (۷) نشان داده شده است.

$$CV = \frac{\sigma}{\mu} \quad (7)$$

در معادله (۷)، μ بیانگر میانگین و σ انحراف معیار می باشد.

۳-۲-۳ ویژگی های پیشرفته

برای مقابله با شرایط ویژه مانند حالات ترافیکی و آب و هوایی، سه ویژگی نرخ تغییر جهت، نرخ توقف و نرخ تغییر سرعت استخراج شده است. مقادیر حد آستانه ویژگی های نرخ تغییر جهت، نرخ توقف و نرخ تغییر سرعت، به ترتیب ۱۹، ۳/۴ و ۰/۲۶ در نظر گرفته شده است (Zheng, Li et al. 2008). در ادامه ویژگی های پیشرفته استخراج شده بررسی می شوند.

- نرخ تغییر جهت: رانندگی با اتومبیل و استفاده از اتوبوس حتی با در نظر گرفتن شرایط ترافیکی و آب و هوایی بارانی، قابلیت تغییر جهت زیادی را نخواهند داشت، در حالی

۳-۳ انتخاب ویژگی

انتخاب ویژگی یک فرآیند استفاده از آزمون‌های آماری برای تعیین ویژگی‌هایی است که موجب بهترین پیش‌بینی شوند، به همین منظور روش‌هایی برای انتخاب ویژگی گسترش داده شده‌اند. استفاده از روش‌های انتخاب ویژگی می‌تواند آموزش را سریع کند، پیچیدگی مدل را کاهش دهد و از بیش‌برازش جلوگیری کند. در این مقاله از یک روش ترکیبی شامل دو روش برمبنای فیلتر^۹ و بر مبنای بسته‌بندی^{۱۰} استفاده شده است (Venkatesh and Anuradha 2019).

– روش برمبنای فیلتر: این روش مانند فیلتر، ویژگی‌هایی را که دارای اهمیت کمتری هستند از مجموعه ویژگی‌ها حذف می‌کند. هنگامی که داده‌های یک ویژگی بیش از حد به هم شبیه باشند، آن ویژگی توانایی پیش‌بینی را نخواهد داشت. به همین منظور با تعیین حد‌آستانه واریانس، ویژگی‌های ثابت و نیمه‌ثابت از داده‌ها حذف خواهد شد. معادله (۱۱) نحوه محاسبه واریانس را نشان می‌دهد، با توجه به این معادله اگر مقادیر مجموعه داده‌ای باهم برابر باشند، واریانس آن صفر محاسبه می‌گردد. همچنین برای حذف ویژگی‌های نیمه ثابت مقدار حد‌آستانه واریانس ۰/۰۱ در نظر گرفته شده است. ویژگی‌هایی که دارای همبستگی زیادی نسبت به یکدیگر هستند علاوه بر تاثیر نداشتن در دقت مدل نهایی، موجب افزایش پیچیدگی مدل یادگیری خواهند شد، به همین منظور پس از تشکیل ماتریس همبستگی، ویژگی‌هایی که مقدار ضریب همبستگی آن‌ها در ماتریس بیش از ۰/۹ می‌باشند نشان‌دهنده همبستگی بسیار زیاد خواهد بود و از مجموعه ویژگی‌ها حذف می‌شوند (Jović, Brkić et al. 2015).

$$\sigma^2 = \frac{\sum_{i=1}^N (x_i - \mu)^2}{N} \quad (11)$$

– روش برمبنای بسته‌بندی: این روش ویژگی‌ها را با توجه به برچسب داده‌ها تنظیم کرده و ویژگی‌هایی که موجب دقت بهینه می‌شوند را انتخاب می‌کند. در این فرآیند از روش جستجوی روبه‌جلو^{۱۱} استفاده شده است (Jović,

Brkić et al. 2015). در این روش پس از انتخاب هر ویژگی، دقت آن با توجه به مدل کلاسه‌بندی انتخابی بررسی شده و ویژگی که بیشترین دقت را داراست انتخاب می‌شود. ویژگی‌های بعدی شامل ویژگی‌های انتخاب شده قبلی نیز می‌باشد. به منظور بالابردن دقت محاسبات، از K-Fold Cross Validation در این روش استفاده شده‌است، در روش K-Fold Cross Validation داده‌های آموزشی به K تعداد تقسیم و بسته‌بندی می‌شوند و فرآیند مورد محاسبه K بار تکرار می‌شود، در هر تکرار یک بسته به عنوان داده اعتبار-سنجی انتخاب شده و تعیین دقت انجام می‌پذیرند. دقت نهایی میانگین K تکرار می‌باشد (Han, Kamber et al. 2011).

۳-۴ کلاسه‌بندی

گام اصلی در تشخیص حالات حمل‌ونقلی کلاسه‌بندی می‌باشد. کلاسه‌بندی به معنای تقسیم داده‌ها براساس شباهتشان است که به صورت نظارت شده پیاده‌سازی شده و قابلیت پیش‌بینی را دارا می‌باشد. کلاس‌های خروجی در این پژوهش شامل: پیاده‌روی، استفاده از دوچرخه، استفاده از اتومبیل، استفاده از اتوبوس و استفاده از قطار می‌باشد که با استفاده از چهار مدل کلاسه‌بندی ماشین بردار پشتیبان، جنگل تصادفی، تقویت کننده گرادیان و تقویت کننده بیش‌ازحد گرادیان انجام می‌پذیرد. در ادامه تعریف مختصری از این روش‌ها ارائه خواهد شد.

۳-۴-۱ ماشین بردار پشتیبان

این مدل با تعیین بردار پشتیبان به بررسی و انتخاب بهترین خط و حاشیه خط می‌پردازد به گونه‌ای که بتوان کلاس‌ها را به بهترین شکل از هم جدا کرد. در صورت پیچیده بودن مدل، تابع خطی توانایی جداسازی را نخواهد داشت و نیاز به تعریف توابع کرنل وجود دارد تا فضای نقاط را تغییر داده و در فضای جدید کلاسه‌بندی انجام گیرد. این مدل از نظر حافظه عملکرد خوبی داشته و برای داده‌های با ابعاد بالا به خوبی کار می‌کند. طبق معادله (۱۲) و حل آن به روش لاگرانژ، کلاسه‌بندی اجرا خواهد شد (Shah 2007).

استخراج ویژگی از خطوط سیر به منظور شناسایی حالت حمل و نقل در شهر هوشمند با استفاده از یادگیری ماشین

به شکل تکراری موجب پیش‌بینی می‌گردد. این مدل سعی در کاهش خطا (بایاس) دارد و همچنین قابلیت تنظیم پارامترها را خواهد داشت که موجب انعطاف پذیری مدل خواهد شد. برخلاف مدل جنگل تصادفی این مدل توانایی آموزش به صورت موازی را نخواهد داشت. معادلات (۱۴) نشان‌دهنده نحوه محاسبه مدل تقویت‌گرایان می‌باشد (Friedman 2001).

$$F_0 = \operatorname{argmin} \sum_{i=1}^n L(y_i, \gamma)$$

$$L(y_i, \gamma) = -\log(\text{likelihood})$$

$$= -[y_i \log(P) + (1 - y_i) \log(1 - P)]$$

$$\frac{d}{d \log(\text{odds})} = -y_i + \frac{e^{\log(\text{odds})}}{1 + e^{\log(\text{odds})}} \quad (14)$$

$$= -\text{observed} + \text{Predicted}$$

$$F_m(x) = F_{m-1}(x) + v \sum_{j=1}^{j_m} \gamma_{jm} I(x \in R_{jm})$$

$$\gamma = \frac{\text{Residual}}{P(1 - P)}$$

در معادلات (۱۴)، L بیانگر تابع هزینه، y_i برچسب مشاهدات، x بیانگر ویژگی‌ها، P احتمال رخداد، v نرخ یادگیری، γ تبدیل چند احتمال یک برگ به یک احتمال، F_0 برآوردگر احتمال مرحله اول و $F_m(x)$ برآوردگر احتمال بقیه مراحل به جز مرحله اول می‌باشد.

۳-۴-۴ تقویت‌کننده بیش‌ازحد گرایان

این مدل بهبود یافته تقویت‌گرایان می‌باشد که می‌تواند با سرعت و دقت بالایی پیش‌بینی را انجام دهد. این مدل با اضافه کردن مقدار منظم‌سازی^{۱۵} موجب کاهش بیش‌برازش شده و به دلیل ارث‌بری از روش جنگل تصادفی توانایی آموزش به صورت موازی را دارد. معادله (۱۵) نحوه محاسبه تابع هزینه در مدل تقویت‌کننده بیش‌ازحد گرایان را نشان می‌دهد (Chen and Guestrin 2016). ترم اول این معادله بیانگر تابع هزینه و ترم دوم منظم‌سازی می‌باشد. T نشان‌دهنده تعداد برگ‌های درخت تصمیم و گاما پالتی تعریف شده توسط کاربر برای هرس درخت است. با حل بسط تیلور تا مرحله دوم، تابع هزینه مطابق معادله

$$\min \frac{1}{2} \|\theta\|^2 + C \sum_{t=1}^m \varepsilon^t \quad (12)$$

s. t. $y^t \theta^T \varphi(x^t) \geq 1 - \varepsilon^t$, $\varepsilon^t \geq 0$
فاصله داده x از مرز تصمیم‌گیری برابر با $|\theta^T x + \theta_0| \geq \rho \|\theta\|$ می‌باشد. طبق معادله (۱۲) به منظور بیشینه کردن حاشیه خط می‌بایست بردار θ کمینه شود. C در این معادله برابر با ضریب جریمه و ε^t خطای حاشیه نرم می‌باشد، همینطور y^t نشان‌دهنده تقسیم‌بندی کلاس است که مقدار آن برای داده‌های بالای خط برابر ۱+ و داده‌های پایین خط برابر با ۱- می‌باشد. برای حل معادله به روش غیر خطی نیاز به یک نگاهت ($x \rightarrow \varphi(x)$) به فضای جدید وجود دارد که با توجه به تعیین تابع کرنل محاسبه می‌گردد. کرنل مورد استفاده در این مسئله کرنل گوسین مطابق با معادله (۱۳) می‌باشد که در آن x^t بیانگر داده‌های آموزشی و x داده‌های جدید می‌باشد.

$$\varphi(x^t)^T \varphi(x) = K(x^t, x)$$

$$= \exp\left(-\frac{\|x^t - x\|^2}{2\sigma^2}\right) \quad (13)$$

۳-۴-۲ جنگل تصادفی

مجموع چند درخت تصمیم جنگل را تشکیل می‌دهد. اگر درختان تصمیم به صورت تصادفی تشکیل شوند، می‌توان جنگل تصادفی ساخت. این مدل یک کلاسه‌بندی ترکیبی می‌باشد که به روش دسته‌بندی^{۱۲} درخت‌های تصمیم را می‌سازد و در پایان با رأی‌گیری، محبوب‌ترین کلاس را تخمین می‌زند. این الگوریتم در سال ۲۰۰۱ گسترش داده شده است و توانسته دقت قابل قبولی را در بین توابع کلاسه‌بندی یادگیری ماشین کسب کند. این مدل سعی در کاهش واریانس دارد و می‌تواند به صورت موازی درخت‌های تصمیم را آموزش دهد. همچنین می‌توان به روش دسته‌بندی درختان تصمیم را ساخت و با رأی‌گیری به کلاسه‌بندی پرداخت (Liaw and Wiener 2002).

۳-۴-۳ تقویت‌گرایان

این مدل در مسائل رگرسیون و کلاسه‌بندی مورد استفاده قرار می‌گیرد، در این روش مدل‌های ضعیف از طریق درختان تصمیم ساخته می‌شود و به صورت تقویت‌کننده^{۱۳} و محاسبه تابع هزینه^{۱۴}

(۱۶) محاسبه می شود.

$$L(y_i, P_i) = \left[\sum_{i=1}^n L(y_i, P_i^m) \right] + \sum_k \Omega(f_k) \quad (15)$$

$$\Omega(f) = \gamma T + \frac{1}{2} \lambda \|w\|^2$$

$$L^t(q) = -\frac{1}{2} \sum_{j=1}^T \frac{(\sum_{i \in I_j} g_i)^2}{\sum_{i \in I_j} h_i + \lambda} + \gamma T \quad (16)$$

که g_i مشتق اول تابع هزینه و h_i مشتق دوم تابع هزینه است.

۳-۵ ارزیابی

پس از پیاده سازی مدل های کلاسه بندی، نیاز به ارزیابی آنها وجود دارد. برای ارزیابی مدل های کلاسه بندی از روش هایی شامل ماتریس پیچیدگی^{۱۶} مطابق با جدول (۲)، Precision و Recall مطابق معادله (۱۷)، FScore مطابق معادله (۱۸) و همچنین از منحنی ROC و محاسبه AUC نیز به منظور نمایش بهتر دقت استفاده شده است (Han, Kamber et al. 2011, Novaković, Veljović et al. 2017).

$$P = \frac{TP}{TP + FP}, R = \frac{TP}{TP + FN} \quad (17)$$

$$F_{Score} = \frac{2PR}{P + R} \quad (18)$$

در ماتریس پیچیدگی ترسیمی در جدول (۲)، مثبت درست نشان دهنده تعداد مواردی است که نمونه عضو کلاس مثبت باشد و به درستی عضو همین کلاس تشخیص داده شود؛ منفی نادرست تعداد مواردی است که نمونه عضو کلاس مثبت باشد و به اشتباه عضو کلاس منفی تشخیص داده شود؛ مثبت نادرست تعداد مواردی است که نمونه عضو کلاس منفی باشد و به اشتباه عضو کلاس مثبت تشخیص داده شود و در نهایت منفی درست تعداد مواردی است که نمونه عضو کلاس منفی باشد و به درستی عضو همین کلاس تشخیص داده شود.

جدول ۲. ماتریس پیچیدگی

کلاس پیش بینی شده		واقعیت
پیش بینی مثبت	پیش بینی منفی	
واقعیت مثبت	مثبت درست (TP)	منفی نادرست (FN)
واقعیت منفی	مثبت نادرست (FP)	منفی درست (TN)

از آنجایی که تعداد کلاس های مورد پیش بینی بیش از دو کلاس می باشد، می بایست نتایج هر کلاس میانگین گیری شود. به همین دلیل از سه میانگین خرد^{۱۷}، کلان^{۱۸} و وزنی^{۱۹} در این پژوهش استفاده شده است. در میانگین کلان مطابق با معادله (۱۹) دقت هر کلاس به تنهایی محاسبه و در نهایت میانگین گیری می شود. در حالت خرد همه کلاس ها با هم در تعیین دقت مشارکت دارند که باتوجه به معادله (۲۰) این مقدار محاسبه می گردد. میانگین وزنی نیز مطابق با معادله (۲۱)، مانند حالت کلان با توجه به وزن هر کلاس، میانگین گیری می شود.

$$P_{macro} = \frac{\sum_{i=1}^n P_i}{N}, R_{macro} = \frac{\sum_{i=1}^n R_i}{N}, F_{Score(macro)} = \frac{2P_{macro}R_{macro}}{P_{macro} + R_{macro}} \quad (19)$$

$$P_{micro} = \frac{\sum_{i=1}^n TP_i}{\sum_{i=1}^n TP_i + \sum_{i=1}^n FP_i}, R_{micro} = \frac{\sum_{i=1}^n TP_i}{\sum_{i=1}^n TP_i + \sum_{i=1}^n FN_i}, F_{Score(micro)} = \frac{2P_{micro}R_{micro}}{P_{micro} + R_{micro}} \quad (20)$$

$$P_{weighted} = \frac{\sum_{i=1}^n P_i \times W_i}{\sum_{i=1}^n W_i}, R_{weighted} = \frac{\sum_{i=1}^n R_i \times W_i}{\sum_{i=1}^n W_i}, F_{Score(weighted)} = \frac{2P_{weighted}R_{weighted}}{P_{weighted} + R_{weighted}} \quad (21)$$

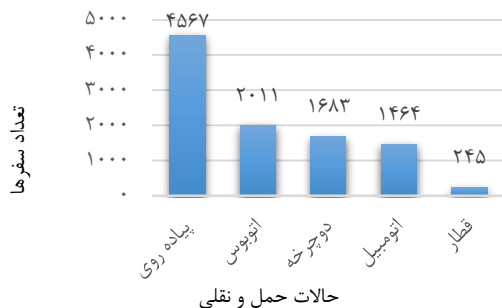
منحنی ROC شامل محور افقی نرخ مثبت نادرست^{۲۰} و محور عمودی نرخ مثبت درست^{۲۱} طبق معادله (۲۲) می باشد. مساحت زیر نمودار منحنی ROC به عنوان یک معیار برای ارزیابی عملکرد کلاسه بندی مورد استفاده قرار می گیرد. در حالت ایده آل، مساحت زیر منحنی برابر با بیشترین مقدار خود، یعنی یک می باشد. هر چه مساحت زیر نمودار به عدد یک نزدیکتر باشد، به معنای بهتر بودن عملکرد کلاسه بندی می باشد (Han, Kamber et al. 2011, Novaković, Veljović et al. 2017).

$$FPR = \frac{FP}{FP + TN}, TPR = \frac{TP}{TP + FN} \quad (22)$$

استخراج ویژگی از خطوط سیر به منظور شناسایی حالت حمل و نقل در شهر هوشمند با استفاده از یادگیری ماشین

۴-۱ پیش پردازش داده‌ها

به دلیل موارد ذکر شده در بخش ۱-۳، نقاطی که بیش از یک بار در مسیر ثبت شده‌اند و دارای مختصات جغرافیایی بیرون از منطقه هستند، حذف می‌شوند. به دلیل اختلالات فرکانسی، سفرهایی که فاصله بین دو نقطه در آن بیشتر از ۱۵ دقیقه می‌باشد به عنوان سفر مجزا در نظر گرفته می‌شود. حد آستانه شتاب و سرعت مطابق با جدول (۱) در داده‌ها اعمال شده و سفرهایی که کمتر از ۳ نقطه داده دارند، حذف می‌شوند، در نهایت سفرهایی که چندین حالات حمل و نقلی در آن استفاده شده، به یک حالت تقسیم‌بندی می‌شوند. شکل (۴) نشان‌دهنده نمودار میله‌ای توزیع داده‌ها می‌باشد، در این شکل پس از انجام پیش-پردازش داده‌ها، ۹۹۷۰ سفر به صورت برجسب‌گذاری ثبت شده است که ۴۵۶۷ سفر پیاده‌روی، ۱۶۸۳ سفر استفاده از دوچرخه، ۲۰۱۱ سفر استفاده از اتوبوس، ۱۴۶۴ سفر استفاده از اتومبیل و ۲۴۵ سفر استفاده از قطار می‌باشد.



شکل ۴. تعداد سفرهای ثبت شده با توجه به حالات حمل و نقلی

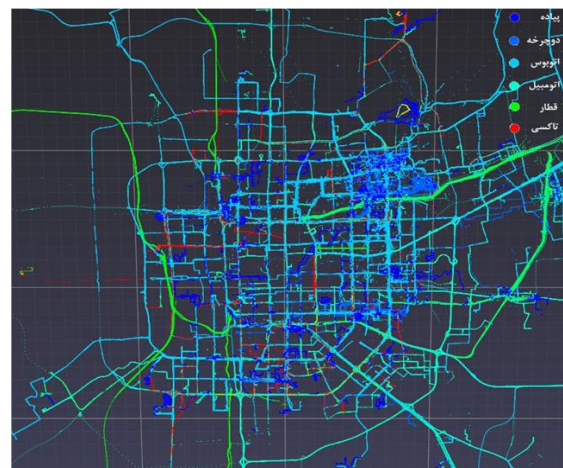
۴-۲ استخراج ویژگی

تمامی ویژگی‌های ذکر شده در بخش ۲-۳ پیاده‌سازی شده است. تعداد چهار ویژگی نقطه‌ای کینماتیک و چهارده ویژگی سفر برای هر یک از ویژگی‌های نقطه‌ای استخراج شده که در مجموع موجب استخراج پنجاه و شش ویژگی سفر شده است. سه ویژگی پیشرفته به منظور بالابردن دقت پیش‌بینی در شرایط ترافیکی و آب‌وهوای بارانی نیز پیاده‌سازی و استخراج شده است. در مجموع پنجاه و نه ویژگی استخراج شده است که نمودار جعبه‌ای

در کلاسه‌بندی داده‌ها با برجسب‌های نامتقارن، استفاده از FScore می‌تواند راه‌حلی به منظور ارزیابی نهایی باشد. در داده‌های چندکلاسی باتوجه به ماهیت برجسب‌ها و معادلات (۲۰، ۲۱، ۲۲)، اگر همه کلاس‌ها دارای یک وزن و برجسب داده نامتقارن باشد استفاده از میانگین خرد در FScore می‌تواند نتیجه نهایی را برآورد کند.

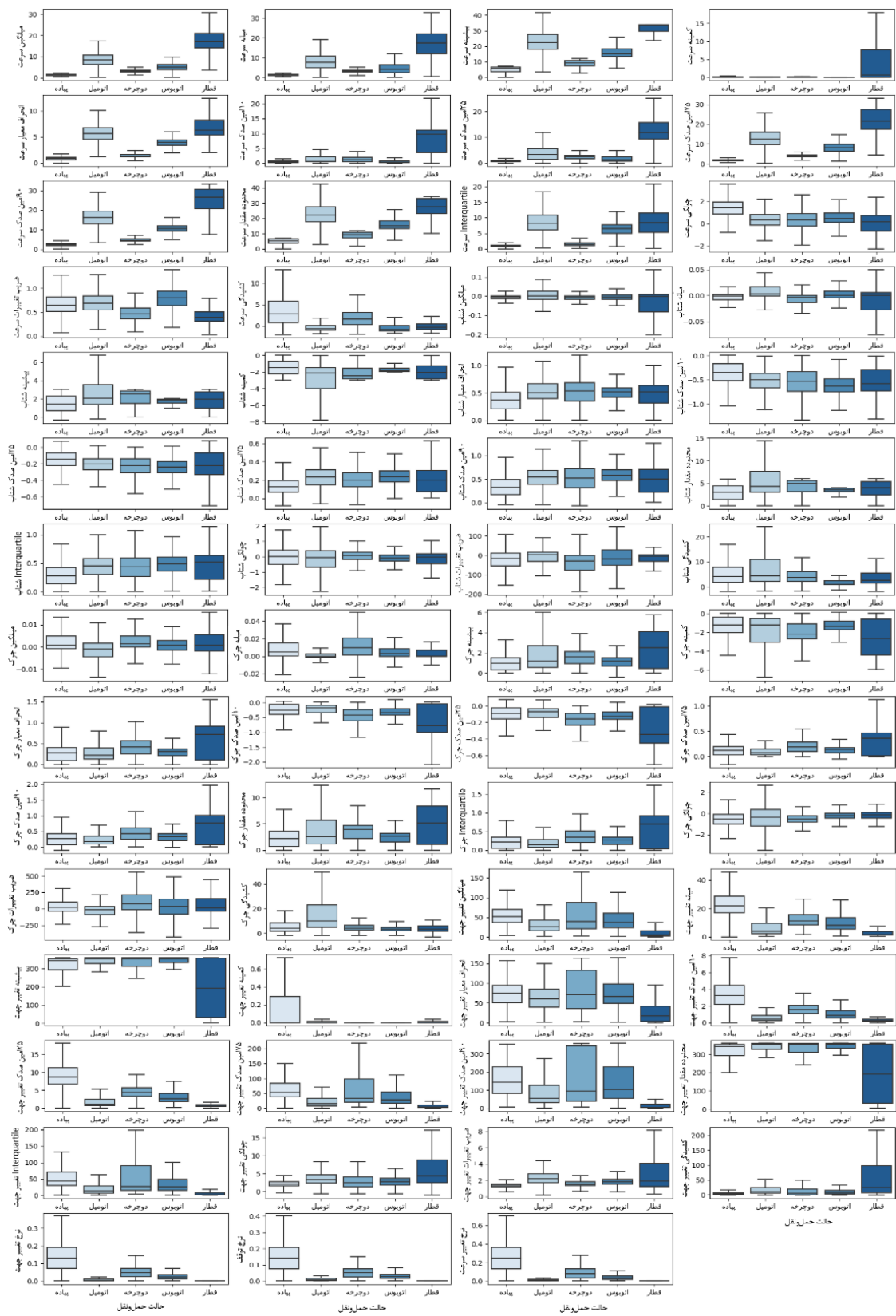
۴. پیاده‌سازی

داده‌های مورد استفاده در این مقاله، مجموعه داده GeoLife است. این مجموعه داده، مسیرهای GPS را در پروژه تحقیقاتی آسیایی مایکروسافت توسط ۱۸۲ کاربر به مدت بیش از ۵ سال (۲۰۰۷-۲۰۱۱) اندازه‌گیری و ذخیره کرده است. این داده‌ها حاوی اطلاعات طول و عرض جغرافیایی، ارتفاع و زمان ثبت می‌باشند. این مجموعه داده توسط دستگاه‌های متنوع ثبت GPS و تلفن‌های همراه هوشمند ذخیره شده‌اند. ۹۱/۵ درصد از مسیرها یک محیط متراکم از نقاط هستند که فاصله ثبت ۱ تا ۵ ثانیه یا ۵ تا ۱۰ متر به ازای هر نقطه را دارند. این مجموعه داده در بیش از ۳۰ شهر چین و در برخی از شهرهای واقع در ایالات متحده آمریکا و اروپا توزیع شده‌اند. حجم بیشتر این داده‌ها در پکن چین ثبت شده‌اند. در این مجموعه داده از بین تمامی کاربران، ۶۹ کاربر داده‌های خود را برجسب‌گذاری کرده‌اند. شکل (۳) نمایشی از این داده‌ها را نشان می‌دهد.



شکل ۳. مجموعه داده GeoLife

آنها در شکل (5) نمایش داده شده است.



شکل 5. نمودار جعبه‌ای و ویژگی‌های استخراج شده

استخراج ویژگی از خطوط سیر به منظور شناسایی حالت حمل و نقل در شهر هوشمند با استفاده از یادگیری ماشین

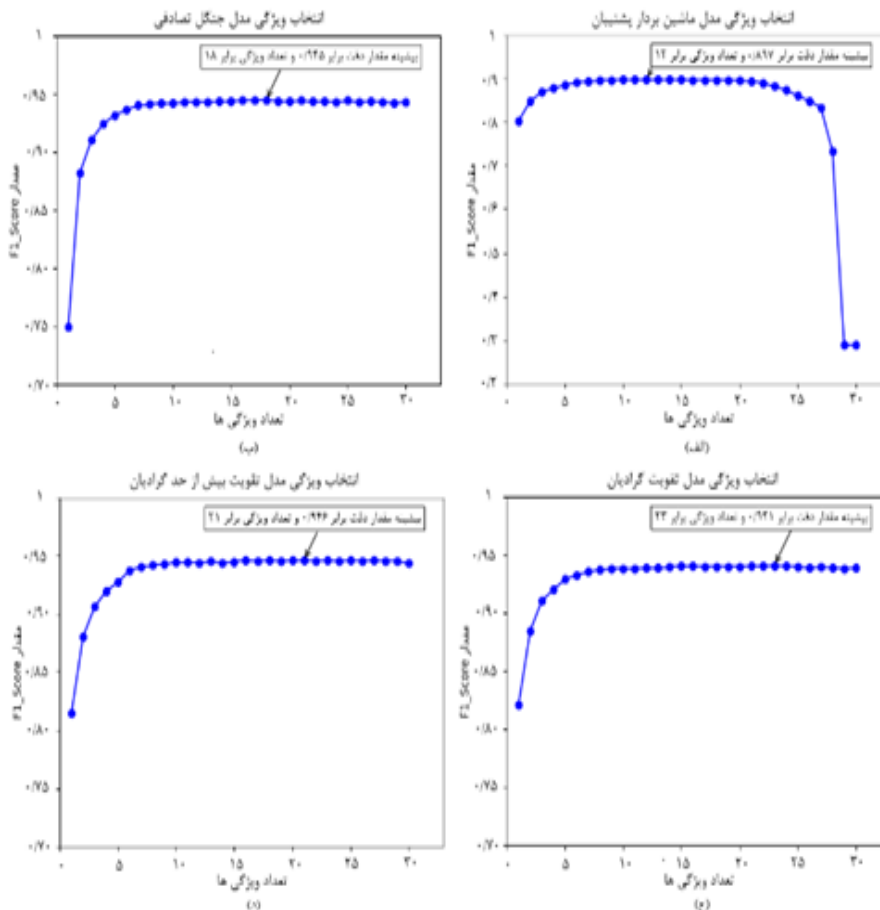
۴-۳ انتخاب ویژگی

برای انتخاب ویژگی، روشی ترکیبی (شامل روش مبتنی بر فیلتر و مبتنی بر بسته‌بندی) روی ویژگی‌های سفر پیاده‌سازی شد. پس از تنظیم و پیاده‌سازی روش مبتنی بر فیلتر، برای هر حالت حمل و نقلی مدل جستجو رو به جلو (یکی از روش‌های مبتنی بر بسته‌بندی)، پیاده‌سازی شد. برای افزایش دقت، برای هر روش کلاسه‌بندی ذکر شده در بخش ۴-۳، انتخاب ویژگی به صورت جداگانه انجام شد. تعداد ویژگی‌های سفر دارای بیشترین دقت به عنوان ویژگی‌های مؤثر در مرحله کلاسه‌بندی بخش ۴-۴، ورودی در نظر گرفته شد. مقدار ورودی پارامترهای کلاسه‌بندی در این مرحله، پیش‌فرض کتابخانه Scikit-learn در نظر گرفته شد. مقادیر دقت، با ۱۰ مرتبه تکرار و محاسبه F_Score به صورت میانگین وزنی با روش 10_Fold Cross Validation، محاسبه شد. پس از پیاده‌سازی، مدل ماشین بردار

پشتیبان با انتخاب ۱۲ ویژگی به دقت ۸۹/۷٪ (شکل ۶-الف)، مدل جنگل تصادفی با انتخاب ۱۸ ویژگی به دقت ۹۴/۵٪ (شکل ۶-ب)، مدل تقویت گرادیان با انتخاب ۲۳ ویژگی به دقت ۹۴/۱٪ (شکل ۶-ج) و مدل تقویت بیش از حد گرادیان با انتخاب ۲۱ ویژگی به دقت ۹۴/۶٪ (شکل ۶-د) دست یافتند.

۴-۴ کلاسه‌بندی

کلاسه‌بندی‌ها با توجه به کتابخانه‌های Scikit-learn و xgboost در زبان برنامه‌نویسی پایتون پیاده‌سازی شد. ۸۰ درصد داده‌ها به عنوان آموزشی و ۲۰ درصد برای تست در نظر گرفته شد. به دلیل نامتعادل بودن برچسب‌ها از یک نمونه‌برداری طبقه‌ای استفاده شده تا بتوان از همه برچسب‌ها در آموزش و تست استفاده کرد. داده‌های آموزشی با استفاده از روش 10_Fold Cross Validation به ۱۰ قسمت تقسیم می‌شود تا بتوان پارامترهای ورودی هر مدل را با بهترین دقت تنظیم کرد.



شکل ۶. تعداد ویژگی‌های انتخاب شده بر اساس مدل کلاسه‌بندی

۴-۱ ماشین بردار پشتیبان

ROC ماشین بردار پشتیبان

جدول ۳. ماتریس پیچیدگی ماشین بردار پشتیبان

کلاس پیش‌بینی شده						
Recall	مجموع	قطار	اتومبیل	اتوبوس	دوچرخه	پیاده روی
۰/۹۹۱	۹۱۳	۰	۱	۱	۶	۹۰۵
۰/۹۰۲	۳۳۷	۰	۰	۸	۳۰۴	۲۵
۰/۸۵۱	۴۰۲	۲	۳۵	۳۴۲	۹	۱۴
۰/۸۴۶	۲۹۳	۳	۲۴۸	۳۶	۱	۵
۰/۶۵۳	۴۹	۳۲	۱۰	۲	۱	۴
-	-	۰/۸۶۵	۰/۸۴۳	۰/۸۷۹	۰/۹۴۷	۰/۹۵۰
-	-	۰/۷۴۴	۰/۸۴۵	۰/۸۶۵	۰/۹۲۴	۰/۹۷۰
						Precision
						F_Score

۴-۲ جنگل تصادفی

در این مدل مقادیر پارامترهای تعداد درخت، کمینه تعداد تقسیم درخت، کمینه تعداد برگ درخت و بیشینه عمق درخت تنظیم شده‌اند. برای هر یک از پارامترها بازه‌ای مطابق با شکل (۸) تعریف می‌شود. برای تعیین هر پارامتر، مقادیر دیگر ثابت در نظر گرفته و از روش 10_Fold Cross Validation و محاسبه FScore استفاده و در نهایت مقدار نهایی به صورت خرد میانگین‌گیری می‌شود. شکل (۸) میزان دقت آموزشی، دقت اعتبارسنجی و واریانس هر یک از اعداد بازه پارامترها را نشان می‌دهد.

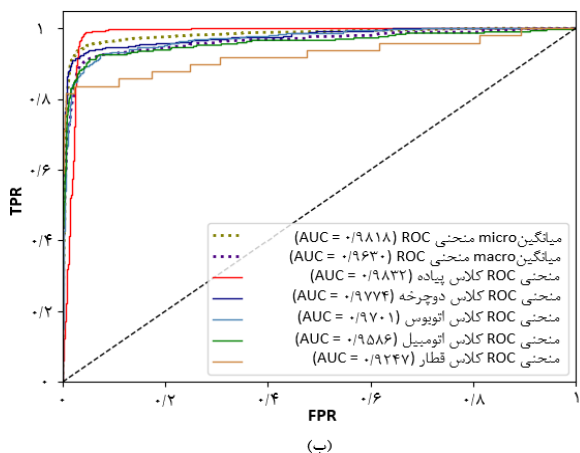
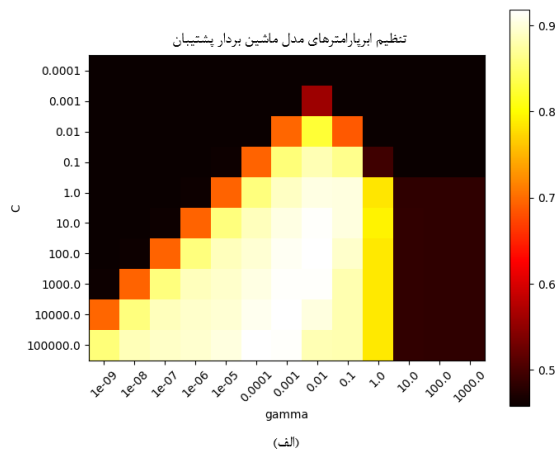
جدول ۴. ماتریس پیچیدگی جنگل تصادفی

کلاس پیش‌بینی شده						
Recall	مجموع	قطار	اتومبیل	اتوبوس	دوچرخه	پیاده
۰/۹۸۷	۹۱۳	۰	۱۰	۱	۹	۹۰۳
۰/۹۴۴	۳۳۷	۰	۰	۵	۳۱۸	۱۴
۰/۹۰۸	۴۰۲	۱	۲۰	۳۶۴	۶	۱۱
۰/۹۰۸	۲۹۳	۰	۲۲۶	۲۱	۱	۵
۰/۷۹۶	۴۹	۳۹	۵	۲	۰	۳
-	-	۰/۹۷۵	۰/۹۱۷	۰/۹۲۹	۰/۹۴۴	۰/۹۶۵
-	-	۰/۸۷۶	۰/۹۱۲	۰/۹۱۸	۰/۹۴۴	۰/۹۷۶
						Precision
						F_Score

پس از انتخاب مقادیری که دارای دقت آموزشی و دقت اعتبار-سنجی بالا و واریانس پایین هستند، از یک پنجره جستجو شامل تمامی مقادیر پارامترهای انتخابی استفاده می‌شود. در پنجره جستجو جدید نیز از روش 10_Fold Cross Validation

فصلنامه مهندسی حمل‌ونقل / سال چهاردهم / شماره اول (۵۴) / پاییز ۱۴۰۱

در ابتدا نیاز است تا پارامترهای C و σ تنظیم شوند. بدین منظور مطابق با شکل (۷-الف) یک بازه برای C و یک بازه برای مقدار گاما تعیین شد. مقدار گاما برابر با $\gamma = \frac{1}{\sigma^2}$ می‌باشد. پس از تعریف بازه برای این دو پارامتر، با استفاده از پنجره جستجو شبکه‌ای به صورت 10_Fold Cross Validation دقت FScore در حالت میانگین خرد بررسی می‌شود. مقدار C = 100 و $\gamma = 0.01$ بالاترین دقت در بازه تعریف شده‌اند که به عنوان پارامتر تنظیم می‌شوند. ماتریس پیچیدگی در جدول (۳) محاسبه شده است، مقدار FScore در حالت میانگین وزنی ۹۱/۷۱٪، در حالت میانگین کلان ۸۶/۹۶٪ و در حالت میانگین خرد ۹۱/۸۲٪ محاسبه شد. طبق نمودار ROC (شکل ۷-ب) میزان AUC در حالت میانگین خرد، ۰/۹۸۱۸ و در حالت میانگین کلان ۰/۹۶۳۰ بدست آمده است.

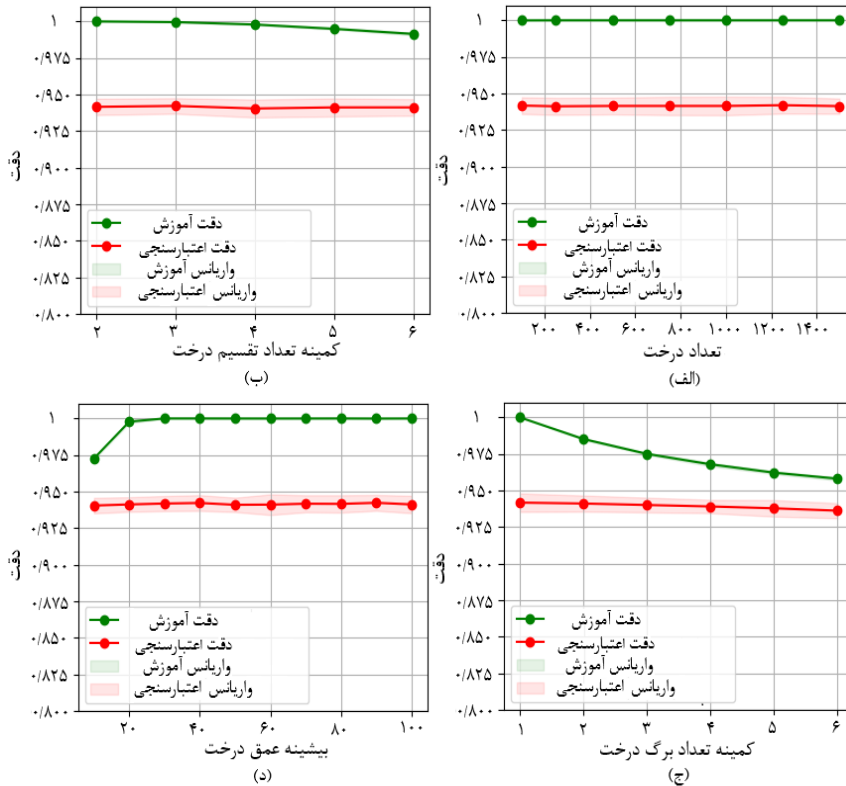


شکل ۷. الف) تنظیم پارامتر ماشین بردار پشتیبان ، ب) نمودار

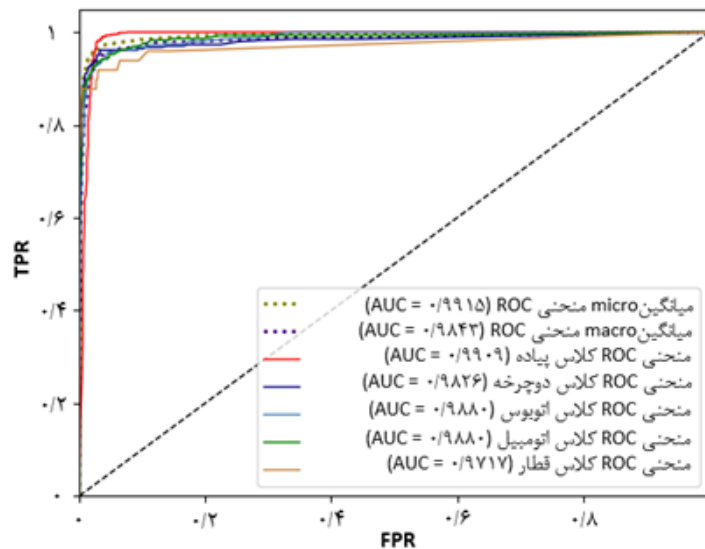
استخراج ویژگی از خطوط سیر به منظور شناسایی حالت حمل و نقل در شهر هوشمند با استفاده از یادگیری ماشین

برابر ۹۴/۶۹ درصد، در حالت میانگین کلان برابر با ۹۲/۵۳ درصد و در حالت میانگین خرد برابر با ۹۴/۷۳ درصد محاسبه شده است. نمودار ROC در شکل (۹) نمایش داده شده، مقدار AUC در حالت میانگین خرد، ۰/۹۹۱۵ و میانگین کلان ۰/۹۸۴۳ بدست آمده است.

و محاسبه FScore به صورت میانگین خرد استفاده می شود. در نهایت مقدار پارامترهای تعداد درخت برابر ۵۰۰، بیشترین عمق برابر ۹۰، کمترین تعداد برگ برابر ۱ و کمترین تعداد تقسیم برابر ۲ به عنوان بهترین مقادیر انتخاب شده است. مطابق ماتریس پیچیدگی جدول (۴)، مقدار FScore در حالت میانگین وزنی



شکل ۸. تنظیم پارامتر مدل جنگل تصادفی



شکل ۹. نمودار ROC مدل جنگل تصادفی

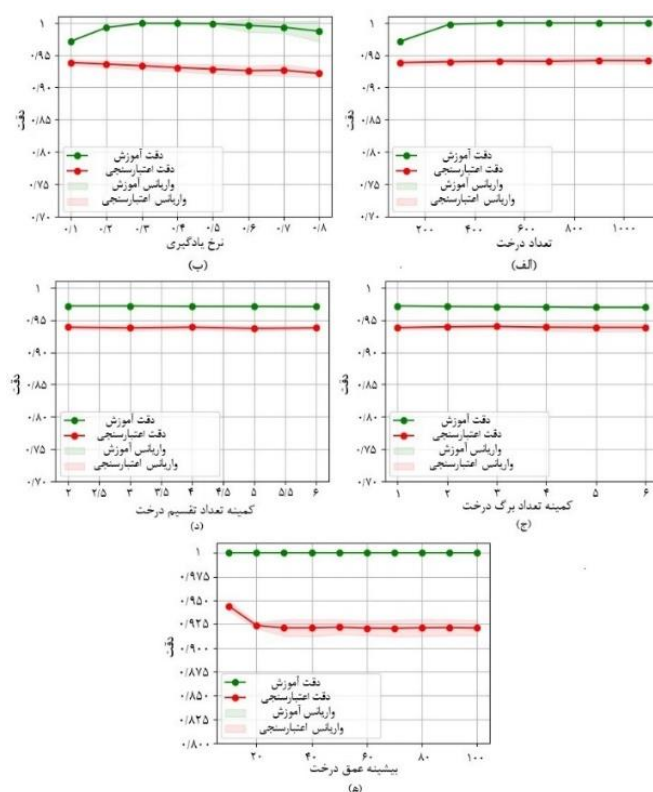
۴-۳ تقویت گرادیان

در این مدل مقادیر پارامترهای تعداد درخت، نرخ یادگیری، کمینه تعداد تقسیم‌بندی درخت، کمینه تعداد برگ و بیشترین عمق درخت تنظیم شده‌اند. بازه‌ای به جهت تنظیم هر یک از پارامترها مطابق با شکل (۱۰) در نظر گرفته شد. هر یک از پارامترها، با ثابت در نظر گرفتن مقادیر دیگر و استفاده از روش 10_Fold Cross Validation، محاسبه F_Score و میانگین‌گیری از آن به صورت خرد تنظیم شده‌اند. شکل (۱۰) نمودار مقادیر دقت داده‌های آموزشی، اعتبارسنجی و واریانس هر یک از مقادیر در نظر گرفته شده پارامترها را نشان می‌دهد. مقادیری که دارای دقت آموزشی و اعتبارسنجی بالا و واریانس کمتری هستند انتخاب می‌شوند و با استفاده از یک پنجره جستجو گریدی تمامی مقادیر پارامترها با هم در نظر گرفته شده و با استفاده از روش 10_Fold Cross Validation و میانگین‌گیری به صورت خرد، مقادیر ۹۰۰ درخت، نرخ یادگیری برابر ۰/۱، کمترین تعداد تقسیم درخت برابر ۲، کمترین تعداد برگ درخت برابر ۳ و بیشترین

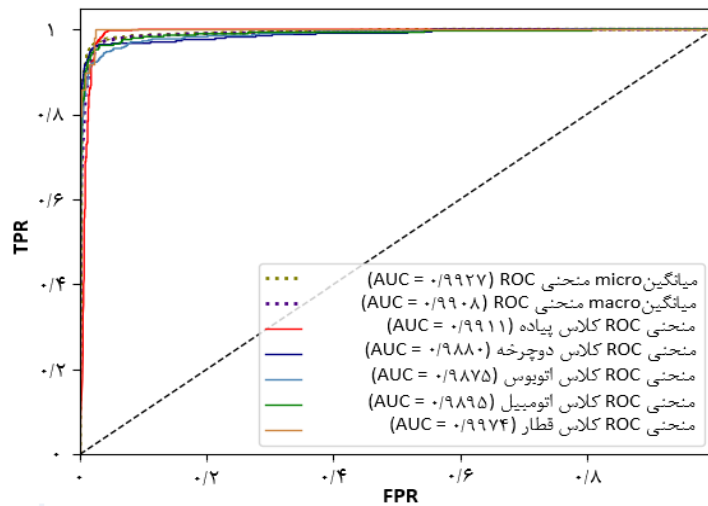
عمق برابر ۱۰ انتخاب شده‌اند. با توجه به ماتریس پیچیدگی جدول (۵)، مقدار F_Score در حالت میانگین وزنی برابر ۹۴/۹۹٪، در حالت میانگین کلان برابر با ۹۲/۳۸٪ و در حالت میانگین خرد برابر با ۹۵/۰۳٪ محاسبه شد. نمودار ROC هر یک از حالات حمل‌ونقلی در شکل (۱۱) نمایش داده شده، مقدار AUC در حالت میانگین خرد، ۰/۹۹۲۷ و میانگین کلان ۰/۹۹۰۸ بدست آمده‌است.

جدول ۵. ماتریس پیچیدگی تقویت گرادیان

کلاس پیش‌بینی شده						
پیاپی	دوچرخه	اتوبوس	اتومبیل	قطار	مجموع	Recall
پیاپی	۵	۴	۰	۰	۹۱۳	۰/۹۹۰
دوچرخه	۳۱۹	۲	۰	۰	۳۳۷	۰/۹۴۶
اتوبوس	۷	۳۶۳	۱۸	۳	۴۰۲	۰/۹۰۳
اتومبیل	۱	۱۶	۲۷۰	۱	۲۹۳	۰/۹۲۱
قطار	۳	۰	۶	۳۹	۴۹	۰/۷۹۶
Precision	۰/۹۶۳	۰/۹۶۱	۰/۹۴۰	۰/۹۱۸	۰/۹۰۷	-
F_Score	۰/۹۷۶	۰/۹۵۴	۰/۹۲۱	۰/۹۲۰	۰/۸۴۸	-



شکل ۱۰. تنظیم پارامتر مدل تقویت گرادیان



شکل ۱۱. نمودار ROC مدل تقویت گرادیان

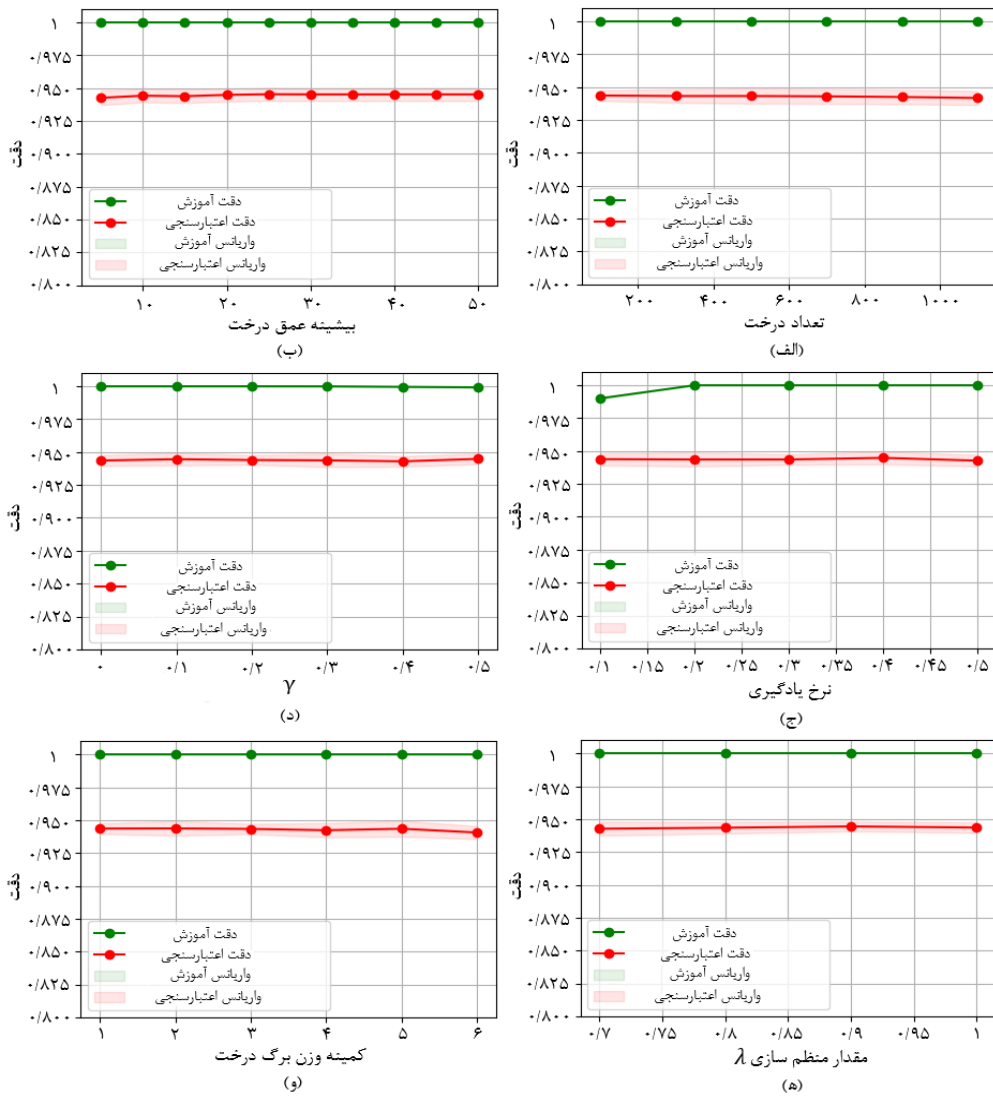
برابر ۰/۵ و منظم‌سازی لامبدا برابر ۰/۹ بالاترین دقت را کسب کرده‌اند. باتوجه به ماتریس پیچیدگی جدول (۶) مقدار FScore در حالت میانگین وزنی برابر ۹۵/۱۴٪، در حالت میانگین کلان برابر با ۹۲/۸٪ و در حالت میانگین خرد برابر با ۹۵/۱۸ محاسبه شده است. نمودار ROC در شکل (۱۳) نمایش داده شده، مقدار AUC در حالت میانگین خرد، ۰/۹۹۴۶ و میانگین کلان برابر با ۰/۹۹۲۱ محاسبه شده است.

جدول ۶. ماتریس پیچیدگی تقویت بیش از حد گرادیان

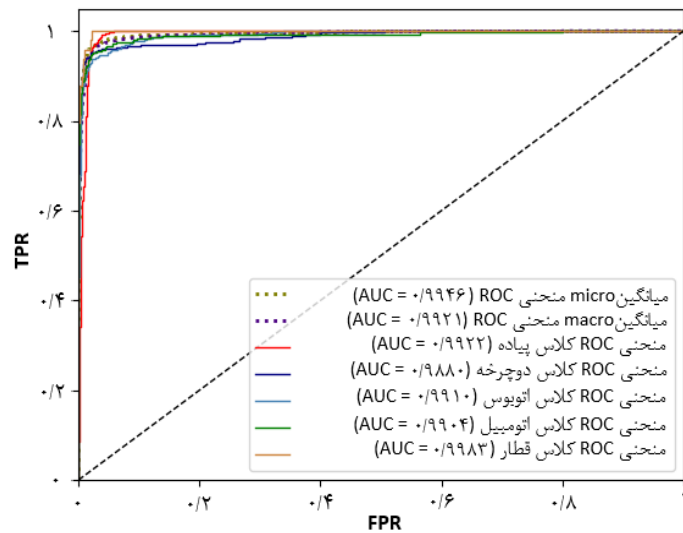
کلاس پیش‌بینی شده							کلاس واقعی
پیاده	دوچرخه	اتوبوس	اتومبیل	قطار	مجموع	Recall	
۹۰۶	۷	۰	۰	۰	۹۱۳	۰/۹۹۲	
۱۵	۳۲۰	۲	۰	۰	۳۳۷	۰/۹۵۰	
۱۱	۸	۳۶۵	۱۶	۲	۴۰۲	۰/۹۰۸	
۵	۲	۱۸	۲۶۸	۰	۲۹۳	۰/۹۱۵	
۳	۰	۲	۵	۳۹	۴۹	۰/۷۹۶	
۰/۹۶۴	۰/۹۵۰	۰/۹۴۳	۰/۹۲۷	۰/۹۵۱	-	Precision	
۰/۹۷۸	۰/۹۵۰	۰/۹۲۵	۰/۹۲۱	۰/۸۶۷	-	F_Score	

۴-۴-۴ تقویت کننده بیش از حد گرادیان

در این مدل مقادیر پارامترهای تعداد درخت، نرخ یادگیری، کمترین وزن برگ درخت، بیشینه عمق درخت، γ و مقدار منظم سازی λ تنظیم شده‌اند. به دلیل چندکلاسه بودن داده‌ها از متد مجموع احتمالات برای هر کلاس استفاده شده است. ابتدا بازه‌ای مطابق با شکل (۱۲) برای هر یک از پارامترها انتخاب می‌شود. هر یک از پارامترها، با ثابت در نظر گرفتن مقادیر دیگر و استفاده از روش 10_Fold Cross Validation، محاسبه FScore و میانگین‌گیری آن به صورت خرد تنظیم شده‌اند. مقادیری که دارای دقت آموزشی و اعتبارسنجی بالا و واریانس کمتری باشند انتخاب و با استفاده از یک پنجره جستجو گریدی به صورت 10_Fold Cross Validation، تمامی مقادیر انتخاب شده پارامترها باهم مورد بررسی قرار گرفته است. با میانگین‌گیری از این مقادیر به صورت خرد، مقادیر ۲۰۰ درخت، نرخ یادگیری برابر ۰/۱، کمترین وزن برگ برابر ۲، بیشترین عمق برابر ۲۵، گاما



شکل ۱۲. تنظیم پارامتر مدل تقویت بیش از حد گرادیان



شکل ۱۳. نمودار ROC مدل تقویت بیش از حد گرادیان

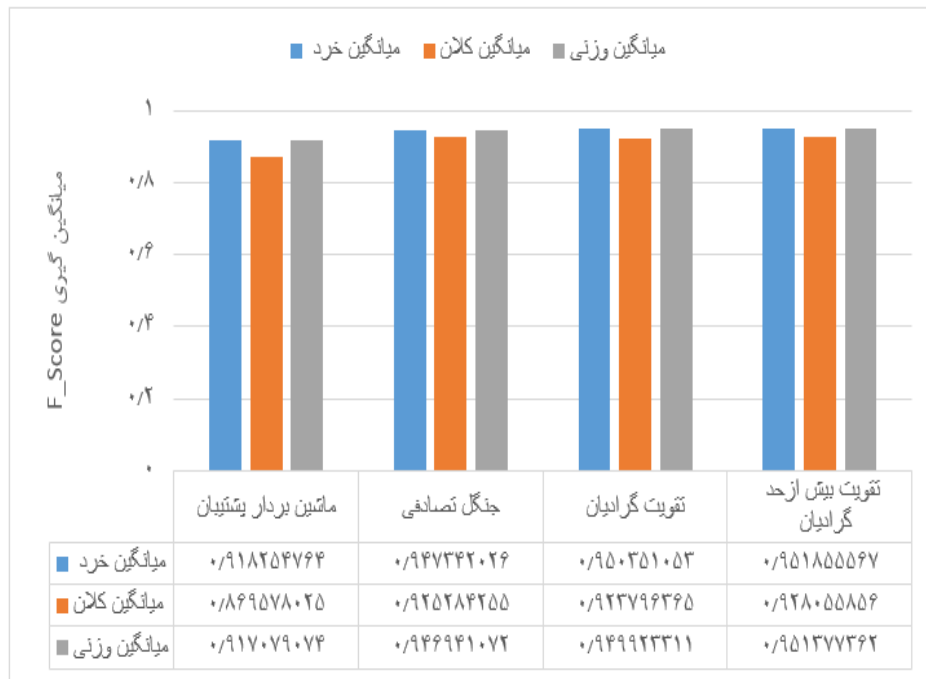
۴-۵ ارزیابی نتایج

در پژوهش‌های قبل، از انتخاب ویژگی ترکیبی برای در نظر گرفتن ویژگی‌های موثر در هر مدل استفاده نشده و همچنین تنظیم پارامترهای هر مدل کلاسه‌بندی به صراحت بیان نگردیده است. در برخی نیز انجام پیش‌پردازش و حذف داده‌های نویزی با رعایت حد آستانه تعیین شده انجام نگرفته است. با توجه به انتخاب ویژگی ترکیبی برای هر مدل کلاسه‌بندی که در بخش ۳-۴ پیاده‌سازی شده است، مدل ماشین بردار پشتیبان با انتخاب ۱۲ ویژگی سفر، مدل جنگل تصادفی با انتخاب ۱۸ ویژگی سفر، مدل تقویت گرادیان با انتخاب ۲۳ ویژگی سفر، مدل تقویت پیش‌ازحد گرادیان با انتخاب ۲۱ ویژگی سفر و همچنین ۳ ویژگی پیش‌رفته برای تمامی مدل‌ها، روش‌های ارزیابی ذکر شده در بخش ۳-۵ در کلاسه‌بندی‌ها پیاده‌سازی شد. با توجه به نامتقارن بودن برچسب داده‌ها، بهترین برآوردگر نهایی دقت، FScore می‌باشد. مقادیر FScore برای تمامی مدل‌های کلاسه‌بندی پیاده‌سازی شده که با توجه به حالات حمل و نقلی در شکل (۱۴) نشان داده شده است. در حالت پیاده، مدل تقویت پیش‌ازحد گرادیان با کسب دقت ۹۷/۷۹ درصد، توانسته بهترین دقت را در بین مدل‌های دیگر بدست آورد. مدل تقویت گرادیان با دقت ۹۵/۳۷ درصد، حالت حمل و نقلی استفاده از دوچرخه را بهتر از مدل‌های دیگر پیش‌بینی کرده است. برای حالت استفاده از اتوبوس، مدل تقویت پیش‌ازحد گرادیان توانسته با کسب دقت ۹۲/۵۲ درصد نسبت به دیگر مدل‌ها، با دقت بالاتری پیش‌بینی این حالت را انجام دهد. در حالت اتومبیل نیز دقت مدل تقویت پیش‌ازحد گرادیان برابر با ۹۲/۰۹ درصد می‌باشد که بهتر از سایر مدل‌ها کلاسه‌بندی این حالت را انجام داده است، همچنین در حالت استفاده از قطار، مدل جنگل تصادفی توانسته با کسب دقت ۸۷/۶۴ درصد بیشترین دقت را کسب کند.

شکل (۱۵) نشان‌دهنده دقت کلی و میانگین‌گیری شده برای تمام حالات با توجه به کلاسه‌بندی‌ها می‌باشد. در این شکل میانگین‌گیری خرد، کلان و وزنی دقت FScore برای تمام حالات حمل و نقلی انجام پذیرفته است. با توجه به هم‌وزن (برتری نداشتن یک کلاس نسبت به کلاس دیگر) و نامتقارن بودن کلاس‌ها، میانگین خرد می‌تواند برآورد نهایی بهتری را داشته باشد. در حالت میانگین خرد، به ترتیب مدل تقویت کننده پیش‌ازحد گرادیان با کسب دقت ۹۵/۱۸ درصد، مدل تقویت گرادیان با کسب دقت ۹۵/۰۳ درصد، مدل جنگل تصادفی با کسب دقت ۹۴/۷۳ درصد و مدل ماشین بردار پشتیبان با کسب دقت ۹۱/۸۲ درصد توانسته‌اند پیش‌بینی کلی را انجام دهند. همچنین در حالت میانگین کلان و وزنی، مدل تقویت پیش‌ازحد گرادیان به ترتیب با کسب دقت ۹۲/۸۰ و ۹۵/۱۳ درصد توانسته دقت بالاتری نسبت به دیگر مدل‌ها کسب کند. به منظور اتکای بیشتر به ارزیابی، شکل (۱۶) نشان‌دهنده میانگین AUC کلاسه‌بندی‌ها می‌باشد. در این شکل مدل تقویت پیش‌ازحد گرادیان با کسب مقدار ۰/۹۹۴۶، مدل تقویت گرادیان با کسب مقدار ۰/۹۹۲۷، مدل جنگل تصادفی با کسب مقدار ۰/۹۹۱۵ و مدل ماشین بردار پشتیبان با کسب مقدار ۰/۹۸۱۸ توانسته‌اند به ترتیب بیشترین مقادیر سطح زیر نمودار ROC را کسب کنند. محاسبات به وسیله سیستم کامپیوتری با اطلاعات 16GB RAM Intel Core i7 4740HQ CPUs و Nvidia Geforce GTX 950M GPUs انجام گرفته است. شکل (۱۷) سرعت آموزش مدل‌های کلاسه‌بندی را با توجه به افزایش داده‌ها نشان می‌دهد. با توجه به شکل (۱۷)، مدل تقویت گرادیان زمان زیادی را صرف آموزش می‌کند و پس از آن به ترتیب مدل جنگل تصادفی، مدل تقویت پیش‌ازحد گرادیان و مدل ماشین بردار پشتیبان زمان بیشتری را به آموزش اختصاص می‌دهند.

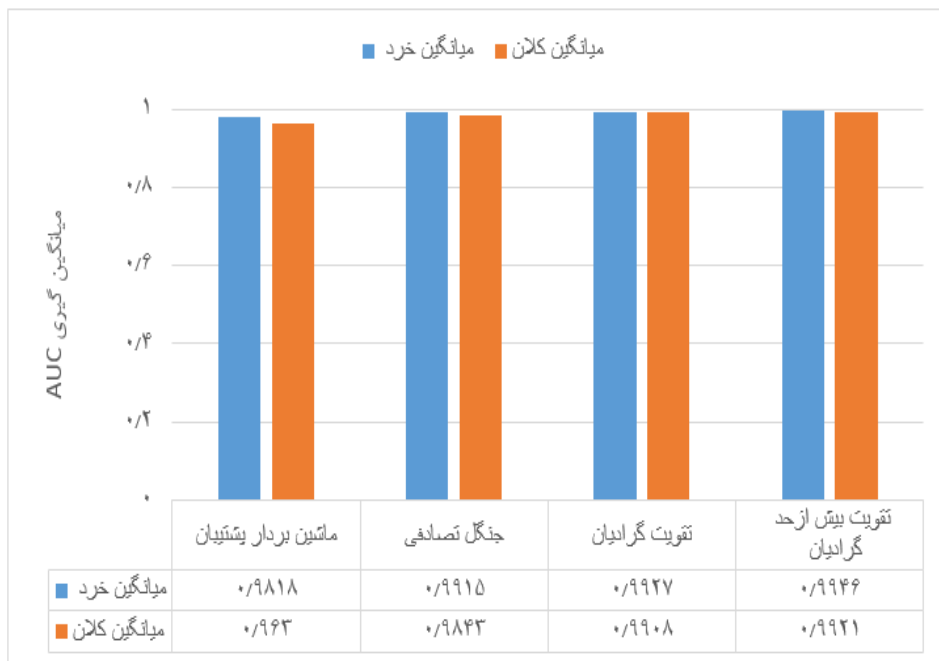


شکل ۱۴. ارزیابی F_Score مدل‌های کلاسه‌بندی

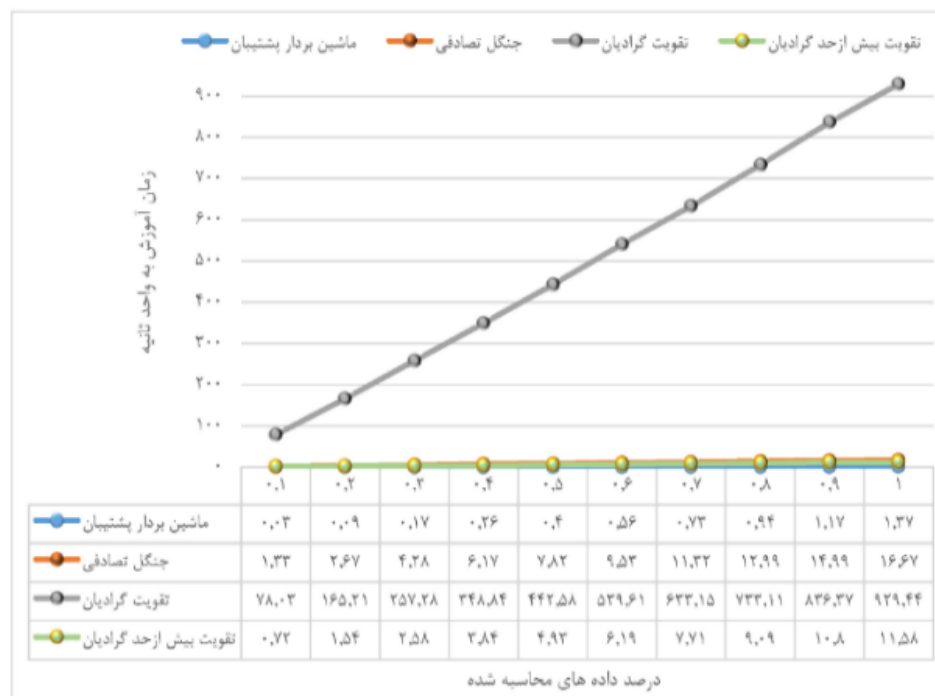


شکل ۱۵. ارزیابی میانگین‌گیری F_score مدل‌های کلاسه‌بندی

استخراج ویژگی از خطوط سیر به منظور شناسایی حالت حمل و نقل در شهر هوشمند با استفاده از یادگیری ماشین



شکل ۱۶. ارزیابی AUC مدل‌های کلاسه‌بندی



شکل ۱۷. سرعت آموزش رویکردهای مختلف

طرف مقابل پیشرفت فناوری امری مهم می‌باشد؛ شهر هوشمند به معنای توسعه شهرها با توجه به پیشرفت فناوری در سطح جهان می‌باشد که اجرای آن موجب افزایش ارزش زندگی افراد می‌گردد. حمل و نقل نقش مهمی در زندگی روزمره شهروندان دارد

۵. نتیجه گیری

عدم توجه به توسعه مراکز شهری موجب شکل‌گیری مشکلات اقتصادی و اجتماعی بزرگی خواهد شد، از این رو نیاز به داشتن شهری هوشمند با توجه به افزایش جمعیت، رشد شهرها و در

فصلنامه مهندسی حمل و نقل / سال چهاردهم / شماره اول (۵۴) / پاییز ۱۴۰۱

، $0.91/0.82$ و مقدار AUC با میانگین خرد $0.99/0.46$ ، $0.99/0.27$ ، $0.99/0.15$ و $0.98/0.18$ کسب کرده‌اند. بهترین مدل از بین مدل‌های بررسی شده مدل تقویت بیش از حد گرادیان می‌باشد که با کسب دقت بهتر ($0.95/0.18$) و پیچیدگی زمانی کمتر توانسته به بهترین شکل پیش‌بینی حالات حمل‌ونقلی را تشخیص دهد. بیشترین خطا میان حالات استفاده از اتومبیل و اتوبوس و همچنین پیاده‌روی و استفاده از دوچرخه در بین مدل‌های پیاده‌سازی شده می‌باشد. استفاده از ویژگی‌های حرکتی و مکمل بیشتر و مدل‌های کلاسه‌بندی دیگر می‌تواند دقت کلی را بهبود بخشد. مطالعات آینده سعی بر پیاده‌سازی مدل شبکه عصبی و استفاده از ویژگی‌های مکمل دیگر به منظور افزایش دقت خواهد بود. همچنین جهت تحقیقات آتی می‌توان بر روی ارزیابی ویژگی‌های مؤثر بر روی حالت‌های گوناگون و تعیین ویژگی‌های مؤثر تمرکز نمود.

۶. پی‌نوشت‌ها

1. Points of Interest (POI)
2. Travel Mode Detection
3. Global Positioning Systems
4. Support Vector Machines (SVM)
5. Random Forest (RF)
6. Gradient Boosting (GB)
7. Extreme Gradient Boosting (XGBoost)
8. Jerk
9. Filter based
10. Wrapper based
11. Forward Search
12. Bagging
13. Boosting
14. Loss Function
15. Regularization
16. Confusion Matrix
17. Micro Average
18. Macro Average
19. Weighted Average
20. False Positive Rate (FPR)
21. True Positive Rate (TPR)

و هر کشور پیشرفته ملزم به توسعه سیستم‌های حمل‌ونقلی خود در جهت تسهیل زندگی افراد می‌باشد. حمل‌نقل هوشمند یکی از زیرمجموعه‌های اصلی شهرهوشمند است که با استفاده از فناوری‌های اطلاعاتی و ارتباطی در زمینه حمل‌ونقل موجب مدیریت زیرساخت‌ها، وسایل نقلیه و جریان‌های ترافیکی خواهد شد؛ به واسطه حمل‌ونقل هوشمند می‌توان با استفاده از فناوری‌های نوین از حداکثر ظرفیت‌های حمل‌ونقلی استفاده نمود. شناسایی حالات حمل‌ونقلی یکی از موارد حمل‌ونقل هوشمند می‌باشد که گام اولیه و مهم برای حل مشکلات حمل‌ونقلی نظیر تحلیل سفر، برنامه‌ریزی حمل‌ونقلی و مدیریت ترافیک به شمار می‌آید. از طرفی گسترش فناوری GPS و تلفن‌های هوشمند توانایی ثبت حجم عظیمی از داده‌ها در کمترین زمان را فراهم آورده و موجب شده تا بتوان سریع‌تر و با دقت‌تر الگوهای پنهان حرکتی را کشف کرد. هدف اصلی این مقاله شناسایی و کشف الگوهای حالات حمل‌ونقلی و پیش‌بینی این حالات با استفاده از داده‌های GPS می‌باشد. استفاده از مختصات داده‌های GPS به تنهایی موجب شناخت حالات حمل‌ونقلی نمی‌گردد و نیازمند استخراج ویژگی‌های مؤثر از مختصات داده‌ها می‌باشد. در این مقاله با استفاده از داده‌های ثبت شده GeoLife و پیش‌پردازش آن، چهار ویژگی نقطه‌ای، چهارده ویژگی سفر برای هریک از ویژگی‌های نقطه‌ای و سه ویژگی پیشرفته استخراج شده است. ویژگی‌های مؤثر سفر با روش ترکیبی انتخاب می‌شوند و به همراه ویژگی‌های پیشرفته به عنوان ورودی وارد مدل‌های کلاسه‌بندی می‌شوند. این پژوهش توانست حالات حمل‌ونقلی پیاده‌روی، استفاده از دوچرخه، استفاده از اتومبیل، استفاده از اتوبوس و استفاده از قطار را با توجه به مدل‌های یادگیری ماشین جنگل تصادفی، ماشین بردار پشتیبان، تقویت گرادیان و تقویت بیش از حد گرادیان پیش‌بینی کند. پس از ارزیابی مدل‌های کلاسه‌بندی، مدل تقویت بیش از حد گرادیان، تقویت گرادیان، جنگل تصادفی و ماشین بردار پشتیبان به ترتیب مقادیر دقت FScore با میانگین خرد $0.95/0.18$ ، $0.95/0.3$ ، $0.94/0.73$

۷. منابع

- Chen, T. and C. Guestrin (2016). Xgboost: A scalable tree boosting system. Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining.
- Dabiri, S. and K. Heaslip (2018). "Inferring transportation modes from GPS trajectories using a convolutional neural network." Transportation research part C: emerging technologies 86: 360-371.
- Dabiri, S. and K. Heaslip (2018). "Transport-domain applications of widely used data sources in the smart transportation: A survey." arXiv preprint arXiv:1803.10902.
- Das, R. D. and S. Winter (2016). "Detecting urban transport modes using a hybrid knowledge driven framework from GPS trajectory." ISPRS International Journal of Geo-Information 5(11): 207.
- Dodge, S., R. Weibel and E. Forootan (2009). "Revealing the physics of movement: Comparing the similarity of movement characteristics of different types of moving objects." Computers, Environment and Urban Systems 33(6): 419-434.
- Endo, Y., H. Toda, K. Nishida and A. Kawanobe (2016). Deep feature extraction from trajectories for transportation mode estimation. Pacific-Asia Conference on Knowledge Discovery and Data Mining, Springer.
- Friedman, J. H. (2001). "Greedy function approximation: a gradient boosting machine." Annals of statistics: 1189-1232.
- Guo, M., S. Liang, L. Zhao and P. Wang (2020). "Transportation Mode Recognition With Deep Forest Based on GPS Data." IEEE Access 8: 150891-150901.
- Adler, J. L. and V. J. Blue (1998). "Toward the design of intelligent traveler information systems." Transportation Research Part C: Emerging Technologies 6(3): 157-172.
- Bagdadi, O. and A. Várhelyi (2013). "Development of a method for detecting jerks in safety critical events." Accident Analysis & Prevention 50: 83-91.
- Bantis, T. and J. Haworth (2017). "Who you are is how you travel: A framework for transportation mode detection using individual and environmental characteristics." Transportation Research Part C: Emerging Technologies 80: 286-309.
- Biljecki, F., H. Ledoux and P. Van Oosterom (2013). "Transportation mode-based segmentation and classification of movement trajectories." International Journal of Geographical Information Science 27(2): 385-407.
- Bolbol, A., T. Cheng, I. Tsapakis and J. Haworth (2012). "Inferring hybrid transportation modes from sparse GPS data using a moving window SVM classification." Computers, Environment and Urban Systems 36(6): 526-537.
- Brennan, C. A., G. Maia, F. Cunha, D. L. Guidoni and L. A. Villas (2019). "Towards a fog-enabled intelligent transportation system to reduce traffic jam." Sensors 19(18): 3916.
- Chamoso, P., A. González-Briones, S. Rodríguez and J. M. Corchado (2018). "Tendencies of technologies and platforms in smart cities: a state-of-the-art review." Wireless Communications and Mobile Computing 2018.

- Langley, R. B. (1997). "Innovation: the GPS error budget." *GPS world* 8(3): 51-56.
- Liaw, A. and M. Wiener (2002). "Classification and regression by randomForest." *R news* 2(3): 18-22.
- Murakami, E. and D. P. Wagner (1999). "Can using global positioning system (GPS) improve trip reporting?" *Transportation research part c: emerging technologies* 7(2-3): 149-165.
- Nawaz, A., H. Zhiqiu, W. Senzhang, Y. Hussain, I. Khan and Z. Khan (2020). "Convolutional LSTM based transportation mode learning from raw GPS trajectories." *IET Intelligent Transport Systems* 14(6): 570-577.
- Novaković, J. D., A. Veljović, S. S. Ilić, Ž. Papić and T. Milica (2017). "Evaluation of classification models in machine learning." *Theory and Applications of Mathematics & Computer Science* 7(1): 39-46-39-46.
- Payne, S. (2015). "Study on key performance indicators for intelligent transport systems: final report in support of the implementation of the EU Legislative Framework on ITS (Directive 2010/40/EU)."
- Quessada, M. S., R. S. Pereira, W. Revejes, B. Sartori, E. N. Gottsfritz, D. D. Lieira, M. A. da Silva, G. P. Rocha Filho and R. I. Meneguette (2020). "ITSMEI: An intelligent transport system for monitoring traffic and event information." *International Journal of Distributed Sensor Networks* 16(10): 1550147720963751.
- Reddy, S., M. Mun, J. Burke, D. Estrin, M. Hansen and M. Srivastava (2010). "Using mobile phones to determine transportation modes." *ACM Transactions on Sensor Networks (TOSN)* 6(2): 1-27.
- Guo, M., P. Wang and L. Zhao (2019). "Research on recognition method of transportation modes based on deep learning." *J. Harbin Inst. Technol.* 51(11): 1-7.
- Han, J., M. Kamber and J. Pei (2011). "Data mining concepts and techniques third edition." *The Morgan Kaufmann Series in Data Management Systems* 5(4): 83-124.
- Huang, Z., P. Wang and Y. Liu (2020). "Statistical characteristics and transportation mode identification of individual trajectories." *International Journal of Modern Physics B* 34(10): 2050092.
- Jahangiri, A. and H. Rakha (2014). "Developing a support vector machine (SVM) classifier for transportation mode identification by using mobile phone sensor data." *Transportation Research Board 93rd Annual Meeting*.
- Jahangiri, A. and H. A. Rakha (2015). "Applying machine learning techniques to transportation mode recognition using mobile phone sensor data." *IEEE transactions on intelligent transportation systems* 16(5): 2406-2417.
- Jarašūniene, A. (2006). "Analysis of possibilities and proposals of intelligent transport system (ITS) implementation in Lithuania." *Transport* 21(4): 245-251.
- Jarašūniene, A. (2007). "Research into intelligent transport systems (ITS) technologies and efficiency." *Transport* 22(2): 61-67.
- Jović, A., K. Brkić and N. Bogunović (2015). "A review of feature selection methods with applications." *2015 38th international convention on information and communication technology, electronics and microelectronics (MIPRO), Ieee.*

- of the existing methodologies." *Information* 7(4): 67.
- Xiao, Z., Y. Wang, K. Fu and F. Wu (2017). "Identifying different transportation modes from trajectory data using tree-based ensemble classifiers." *ISPRS International Journal of Geo-Information* 6(2): 57.
- Yang, X., K. Stewart, L. Tang, Z. Xie and Q. Li (2018). "A review of GPS trajectories classification based on transportation mode." *Sensors* 18(11): 3741.
- Yu, J. (2019). "Travel mode identification with gps trajectories using wavelet transform and deep learning." *IEEE Transactions on Intelligent Transportation Systems*: 1-11.
- Zheng, Y., Y. Chen, Q. Li, X. Xie and W.-Y. Ma (2010). "Understanding transportation modes based on GPS data for web applications." *ACM Transactions on the Web (TWEB)* 4(1): 1-36.
- Zheng, Y., H. Fu, X. Xie, W.-Y. Ma and Q. Li (2011). "Geolife GPS trajectory dataset-user guide." Microsoft Research.
- Zheng, Y., Q. Li, Y. Chen, X. Xie and W.-Y. Ma (2008). Understanding mobility based on GPS data. *Proceedings of the 10th international conference on Ubiquitous computing*.
- Zheng, Y., L. Liu, L. Wang and X. Xie (2008). Learning transportation mode from raw gps data for geographic applications on the web. *Proceedings of the 17th international conference on World Wide Web*.
- Zhong, R., A. Sumalee, T. Pan and W. Lam (2013). "Stochastic cell transmission model for traffic network with demand and supply uncertainties." *Transportmetrica A: Transport Science* 9(7): 567-602.
- Scheiner, J. and C. Holz-Rau (2007). "Travel mode choice: affected by objective or subjective determinants?" *Transportation* 34(4): 487-511.
- Shah, R. S. (2007). *Support vector machines for classification and regression*, McGill University.
- Song, X., H. Kanasugi and R. Shibasaki (2016). Deeptransport: Prediction and simulation of human mobility and transportation mode at a citywide level. *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*.
- Stenneth, L., O. Wolfson, P. S. Yu and B. Xu (2011). Transportation mode detection using mobile phones and GIS information. *Proceedings of the 19th ACM SIGSPATIAL international conference on advances in geographic information systems*.
- Tamane, S. C., N. Dey and A. E. Hassani (2020). *Security and Privacy Applications for Smart City Development*, Springer.
- Venkatesh, B. and J. Anuradha (2019). "A review of feature selection and its methods." *Cybernetics and Information Technologies* 19(1): 3-26.
- Vincnty, T. (1975). "Direct and inverse solutions of geodesics on the ellipsoid with application of nested equations." *Survey Review*.
- Vu, T. H., L. Dung and J.-C. Wang (2016). Transportation mode detection on mobile devices using recurrent nets. *Proceedings of the 24th ACM international conference on Multimedia*.
- Wu, L., B. Yang and P. Jing (2016). "Travel mode detection based on GPS raw data collected by smartphones: a systematic review

- Zhu, Q., M. Zhu, M. Li, M. Fu, Z. Huang, Q. Gan and Z. Zhou (2016). Identifying transportation modes from raw GPS data. International Conference of Pioneering Computer Scientists, Engineers and Educators, Springer.

استخراج ویژگی از خطوط سیر به منظور شناسایی حالت حمل و نقل در شهر هوشمند با استفاده از یادگیری ماشین

رحیم علی عباسپور، درجه کارشناسی در رشته مهندسی نقشه برداری را در سال ۱۳۷۹ از دانشگاه تهران و درجه کارشناسی ارشد در رشته مهندسی نقشه برداری - سیستم‌های اطلاعات مکانی را در سال ۱۳۸۱ از دانشگاه تهران اخذ نمود. در سال ۱۳۸۹ موفق به کسب درجه دکتری سیستم‌های اطلاعات مکانی از دانشگاه تهران گردید. زمینه‌های پژوهشی مورد علاقه ایشان داده کاوی مکانی-زمانی، بهینه‌سازی مکانی، اطلاعات مکانی داوطلبانه و خدمات مکان‌مبنا (LBS) بوده و در حال حاضر عضو هیات علمی با مرتبه دانشیار در پردیس دانشکده‌های فنی دانشگاه تهران است.



علیرضا چهرقان در سال ۱۳۸۸ رشته کارشناسی مهندسی نقشه برداری را از دانشگاه تهران اخذ نمود. همچنین کارشناسی ارشد و دکتری خود را به ترتیب در سال‌های ۱۳۹۰ و ۱۳۹۶ از دانشگاه تهران دریافت کرد. زمینه‌های پژوهشی مورد علاقه ایشان شامل تناطریابی عوارض در پایگاه‌های داده مکانی، اطلاعات مکانی داوطلبانه، داده کاوی مکانی-زمانی، تصمیم‌گیری‌های مکان‌مبنا، بهینه‌سازی مکانی، خدمات مکان‌مبنا و محاسبات هندسی می‌باشد. در حال حاضر ایشان عضو هیات علمی با مرتبه استادیار در دانشگاه صنعتی سهند است.



سجاد صولتی، درجه کارشناسی در رشته مهندسی عمران-نقشه‌برداری را در سال ۱۳۹۸ با رتبه ۵ درصد برتر از دانشگاه تهران به پایان رسانده و هم‌اکنون دانشجوی کارشناسی ارشد مهندسی سیستم‌های اطلاعات مکانی می‌باشد. زمینه‌های پژوهشی مورد علاقه ایشان تجزیه و تحلیل داده‌های مکانی، بررسی کلان‌داده‌های مکانی و پیش‌بینی شناسایی حالات‌های حمل و نقل می‌باشد.

