فصلنامه مهندسی حمل و نقل، دوره ۱۴، شماره ۳، بهار ۱۴۰۲، ص. ۲۷۴۹-۲۷۲۷
مقاله علمی-پژوهشی

مطالعه رفتار سفر افراد با استفاده از داده‌های تلفن همراه (CDR)

مطالعه موردی: شهر تهران

مهدی‌یار درویش‌زاده کاخکی، دانشجوی کارشناسی ارشد، گروه مهندسی و برنامه‌ریزی حمل‌ونقل، دانشکده مهندسی عمران، دانشگاه علم و صنعت ایران، ایران
سعید رحمانی، کارشناس ارشد، مرکز تحلیل داده‌های حمل‌ونقل، دانشگاه علم و صنعت ایران، ایران
افشین شریعت (مسئول مکاتبات)، استاد، گروه برنامه‌ریزی حمل‌ونقل، دانشکده مهندسی عمران، دانشگاه علم و صنعت ایران، ایران
E-mail: shariat@iust.ac.ir
سینا عبداللهی، دانشجوی کارشناسی ارشد، گروه مهندسی و برنامه‌ریزی حمل‌ونقل، دانشکده مهندسی عمران، دانشگاه علم و صنعت ایران، ایران

دریافت: ۱۴۰۰/۰۸/۱۱ پذیرش: ۱۴۰۱/۰۱/۲۸

چکیده

پژوهش حاضر سعی بر شناسایی رفتار سفر افراد در شهر تهران با استفاده از داده‌های تلفن همراه (CDR)، نشان دادن قابلیت‌ها و اهمیت استفاده از این نوع داده‌ها در مقایسه با داده‌های سنتی پرسشگری خانوار، دارد. در این مطالعه با استفاده از روش‌شناسی ارائه‌شده سعی می‌شود کارآیی و اهمیت استفاده از این داده‌ها مشخص گردد. این نوع داده‌ها گرچه هزینه کم‌تری را نسبت به سایر داده‌های سنتی بر مطالعه تحمیل می‌کنند، تفکیک‌پذیری زمانی و مکانی کمی نسبت به سایر داده‌های تلفن همراه (مانند GPS) دارند و کاربرد آن‌ها در ابعاد کلان مورد تردید است. در این مطالعه سعی می‌شود اهمیت و کاربرد این داده‌ها اثبات گردد. بدین منظور، در ابتدا با استفاده از روش‌های کار با کلان‌داده‌ها و معیارهای زمانی و مکانی مختلف، داده‌ها پالایش شده و به نمونه‌ای با حدود ۴۰۰ هزار کاربر با ۵۰۰ میلیون رکورد ارتباطی، تقلیل یافت. سپس با استفاده از روش‌های مختلف یادگیری ماشین، به شناسایی نقاط توقف، فعالیت، و اهداف سفر افراد پرداخته شد. با شناسایی سفرهای افراد، ماتریس مبدا-مقصد برای نمونه برآورد شد و با استفاده از ضرایب تعمیم محاسبه‌شده، به مناطق ۲۲گانه شهر تهران تعمیم داده شد. مشاهده شد که این ماتریس تطابق بالایی با الگوهای سفر افراد در شهر تهران که از مطالعات جامع حمل‌ونقل شهر تهران به دست آمده است، دارد. اعتبارسنجی نتایج با مطالعات طرح جامع نشان می‌دهد که روش به‌کارگرفته‌شده در این پژوهش می‌تواند در مقاصد مختلف برنامه‌ریزی حمل‌ونقل بکار گرفته شود.

واژه‌های کلیدی: الگوی سفر، داده‌های تلفن همراه، رفتار سفر، ماتریس مبدا-مقصد، کلان‌داده

۱. مقدمه و پیشینه پژوهش

شناسایی رفتار سفر افراد از اهمیت بالایی در بهبود زیرساخت‌های حمل‌ونقلی، دسترسی افراد به تسهیلات رفاهی و در نهایت بهبود کیفیت زندگی شهروندان، برخوردار است. درگذشته این مهم از طریق آماربرداری‌ها و پرسشگری‌های مختلف، مانند پرسشگری‌های خانوار در سطح شهر، استان، یا کشور، صورت می‌گرفته است. این آماربرداری‌ها همچنان از ابزارهای محبوب و پرکاربرد در مطالعات برنامه‌ریزی حمل‌ونقل محسوب می‌شوند. بااین‌حال، به دلیل هزینه بالای آماربرداری‌های سنتی در اغلب موارد نمونه مورداستفاده در این مطالعات در مقایسه با جمعیت جامعه کوچک بوده و تنها درصد کمی (حدود ۱ الی ۲ درصد) از شهروندان مورد پرسشگری قرار می‌گیرند. همچنین این برداشت‌ها به طور معمول با فواصل زمانی طولانی صورت گرفته و به دلیل تغییرات پیوسته رفتار سفر در کلان‌شهرها با گذشت زمان، این داده‌ها دیگر نماینده خوبی برای رفتار سفر همه کاربران نخواهند بود [Board et al., 2018]. از سوی دیگر، این پرسشگری‌ها معمولاً در روزهای خاصی از سال انجام می‌گیرند و رفتار سفر افراد در این روزها به کل سال تعمیم داده می‌شود [C. Chen et al., 2016]. بدیهی است علی‌رغم استفاده از روش‌های آماری برای کاهش خطا، این‌گونه تعمیم‌ها ناگزیر با خطا همراه خواهند بود. همچنین، خروجی این نوع از پرسشگری‌های محدود، در روزهای خاصی از سال که رفتار سفر از الگوهای گذشته پیروی نمی‌کنند، قابل‌استفاده نمی‌باشد و در صورت نیاز به شناسایی الگوهای جابجایی در روزهای خاص، لازم است مجدداً عملیات آماربرداری تکرار شود. نکته منفی دیگر در خصوص داده‌های پرسشگری خانوار، عدم شناسایی برخی از سفرهای خاص کاربران در طول روز است؛ از جمله عدم گزارش برخی سفرهای کم‌اهمیت از نگاه کاربران مانند سفرهای تفریحی، خرید، و غیره [Board et al., 2018]. این سفرها، بخصوص در روزها و ساعات غیرکاری، حائز اهمیت خواهند بود. این موارد سبب شده‌اند که محققین در سال‌های اخیر به استفاده از روش‌های جایگزین یا مکمل، کم‌هزینه و قابل اعتماد، برای جبران یا کاهش این محدودیت‌ها روی آورند. یکی از این روش‌ها، استفاده از داده‌هایی است که توسط سازمان‌ها و شرکت‌های مختلف، برای مقاصدی غیر از مقاصد حمل‌ونقلی جمع‌آوری می‌شوند، اما می‌توان از آن‌ها به‌عنوان داده‌های مکمل، و یا در برخی موارد، جایگزین در مطالعات حمل‌ونقل و مدیریت شهری استفاده نمود. ازجمله این داده‌ها، می‌توان به داده‌های سلولی تلفن‌های همراه، داده‌های GPS دستگاه‌های ارتباطی مختلف، ازجمله تلفن‌های همراه و خودروهای مجهز، داده‌های بلوتوث و سایر داده‌های مبتنی بر دستگاه‌های قابل‌حمل اشاره نمود. در این میان، داده‌های سلولی تلفن همراه از محبوبیت بالایی در تحلیل رفتار سفر کاربران برخوردار هستند و در سال‌های اخیر در بسیاری از کلان‌شهرهای دنیا برای تحلیل رفتار سفر شهروندان و سایر تحلیل‌های حمل‌ونقلی، مورداستفاده قرار گرفته‌اند [Board et al., 2018; Imai et al., 2021]. افزایش ضریب نفوذ تلفن‌های همراه در کنار پیشرفت‌های ایجادشده در ابزارها و روش‌های تحلیل داده، این وسایل ارتباط جمعی را به حسگرهای ارزشمندی برای جمع‌آوری داده‌های کاربران و مطالعه رفتار سفر آن‌ها، بدل نموده است. البته همواره چالش‌هایی بر سر راه استفاده تحقیقاتی از این نوع داده‌ها بوده است. از جمله این چالش‌ها می‌توان به ماهیت کلان این نوع داده‌ها اشاره کرد که تحلیل آن‌ها نیازمند استفاده از روش‌های نوین و نسبتاً پیچیده است. از سوی دیگر، حفظ حریم شخصی کاربران تلفن همراه همواره یکی از نگرانی‌های مطرح در زمینه استفاده از این نوع داده‌ها می‌باشد [C. Wang & Hess, 2021]. در مطالعه پیش‌رو نیز امکان استفاده از داده‌های تلفن همراه با در نظر گرفتن مسئله حفظ حریم شخصی افراد بررسی گردیده است.

داده‌های تلفن همراه را می‌توان به چهار دسته مختلف تقسیم کرد. اولین نوع از این داده‌ها داده‌های ثبت جزئیات تماس (CDR[1]) هستند که در مطالعه پیش‌رو مورد استفاده قرار

گرفته‌اند. داده‌های ثبت جزئیات تماس شامل اطلاعات موقعیت تماس، پیامک، یا اتصال به اینترنت افراد هستند. این موقعیت در قالب طول و عرض جغرافیایی هر رکورد ثبت می‌شود. داده‌های ثبت جزئیات تماس به صورت پیوسته توسط اپراتورهای تلفن همراه جهت محاسبه کارکرد مشترکین جمع‌آوری شده، که این ویژگی، این نوع از داده‌ها را ارزان و مناسب برای مطالعات بلندمدت کرده است. همچنین به منظور حفظ حریم شخصی افراد، شماره تلفن ایشان رمزگذاری شده است. داده‌های ثبت جزئیات تماس برحسب تفکیک‌پذیری مکانی، به دو نوع کلی آنتن‌مبنا[2] و مثلث‌سازی‌شده[3]، تقسیم‌بندی می‌شوند. داده‌های آنتن‌مبنا صرفاً شامل موقعیت آنتنی است که فرد به آن متصل شده است. اما در داده‌های مثلث‌سازی‌شده، با استفاده از الگوریتم‌های پیشرفته بر مبنای آنتن‌های مجاور افراد در هر زمان، و با اندازه‌گیری قدرت سیگنال ارتباطی، موقعیت فرد تخمین زده می‌شود. تفکیک‌پذیری مکانی این داده‌ها از داده‌های آنتن‌مبنا بیشتر است، اما داده‌های آنتن‌مبنا از دقت کافی برای استفاده در مطالعات حمل‌ونقلی برخوردار هستند [Board et al., 2018]. از سوی دیگر، استفاده از داده‌های ثبت جزئیات تماس نیاز به توسعه زیرساخت‌های جدید و صرف هزینه‌های بالا ندارد و توسط اپراتورهای مخابراتی به‌منظور محاسبه میزان کارکرد مشتریان، جمع‌آوری می‌شوند. این داده‌ها به دلیل پیوستگی جمع‌آوری در روزهای مختلف سال، اطلاعات مربوط به الگوی سفر شهروندان را با جزئیات بیشتری در اختیار تحلیل‌گران قرار می‌دهند. نکته مثبت دیگر در خصوص داده‌های ثبت جزئیات تماس این است که با توجه به ضریب نفوذ بالای تلفن‌های همراه در بین کاربران، اندازه نمونه قابل دست‌یابی در مقایسه با پرسشگری‌های سنتی، عموماً بالاتر بوده و به‌تبع آن، قابلیت اطمینان نتایج حاصل از تحلیل این نوع از داده‌ها عمدتاً بیش‌تر از داده‌های پرسشگری خانوار است. این مزایا سبب شده است که در بسیاری از کشورها از داده‌های ثبت جزئیات تماس در حوزه‌های مختلف، به‌منظور تحلیل الگوی سفر کاربران استفاده شود [Board et al., 2018].

نوع دیگری از داده‌های تلفن همراه داده‌های سیگنالینگ تلفن همراه[4] هستند که علاوه بر رویدادهای ارتباطی افراد (تماس، پیامک، و اتصال به اینترنت) رویدادهای handover، LAU[5]، attachment، و detachment را نیز شامل می‌شوند. داده‌های به‌روزرسانی موقعیت دوره‌ای (PLU[6]) نوع دیگری از داده‌های تلفن همراه هستند. در این نوع داده‌ها، موقعیت افراد به صورت دوره‌ای و پیوسته به‌روزرسانی می‌گردد. استفاده از این داده‌ها نیازمند همکاری اپراتورهای تلفن همراه و هزینه اضافی است. داده‌های مبتنی بر اپلیکیشن‌های هوشمند نیز در سال‌های اخیر مورد استفاده قرار گرفته‌اند که مستلزم نصب اپلیکیشن بر روی تلفن همراه توسط افراد است.

ایده استفاده از داده‌های تلفن همراه به عنوان یکی از ابزارهای شناسایی رفتار سفر افراد از سال ۱۹۹۹ مطرح شد [Calabrese, Colonna, et al., 2011]. درابتدا استفاده از این نوع داده‌ها بصورت کلان، به دلیل محدودیت‌های سخت افزاری، امکان‌پذیر نبود. به تدریج با گسترش توان پردازشی رایانه‌ها، این داده‌ها بسیار مورد توجه قرار گرفتند. تا کنون از داده‌های تلفن همراه در حوزه‌های شناسایی رفتار سفر، نظیر شناسایی نقاط توقف افراد [Vanhoof et al., 2018]، شناسایی نوع فعالیت افراد و اهداف سفر، ارزیابی انتخاب مسیر توسط افراد [Sakamanee et al., 2020]، شناسایی شیوه سفر [Chin et al., 2019]، ساخت مدل های آماری رفتار سفر افراد [Schneider et al., 2013]، ساخت و پیش‌بینی زنجیره سفر افراد [G. Chen et al., 2019; Hoteit et al., 2014]، ساخت ماتریس‌های مبدا–مقصد [Mamei et al., 2019]، مطالعات زمان سفر [Janecek et al., 2012]، شبیه‌سازی در لحظه وضعیت ترافیکی شبکه [Calabrese, Colonna, et al., 2011; Lwin et al., 2018]، شناسایی، دسته‌بندی و میزان استفاده از مسیرها در سطح شهر [Toole et

al., 2015; P. Wang et al., 2012]، استفاده شده است. علاوه بر این، از داده‌های تلفن همراه در حوزه‌هایی جز حمل و نقل، نظیر مطالعات جابجایی جمعیت [Hankaew et al., 2019]، توریسم [Qin et al., 2019; Sikder et al., 2016]، مدیریت بحران [Ghurye et al., 2016]، ارزیابی اقتصاد شهری [Hernandez et al., 2017]، و سلامت، نظیر چگونگی شیوع بیماری‌های واگیردار (مانند COVID-19) [Grantz et al., 2020] و تاثیر آن بر الگوهای حرکتی افراد [Cintia et al., 2020] استفاده شده است. در ادامه به تفصیل به مرور ادبیات موضوع خواهیم پرداخت.

کاسرس و همکاران [Caceres et al., 2008]، مطالعه‌ای را پیرامون امکان‌سنجی ساخت ماتریس مبدا–مقصد و شمارش حجم ترافیک توسط داده‌های تلفن همراه انجام دادند. در این مطالعه با توجه به ماهیت غیرفعال داده‌های تلفن همراه، از یک پایگاه داده شبیه‌سازی‌شده داده‌های تلفن همراه با استفاده از داده های واقعی یک روزه استفاده کردند که فقط شامل تعداد تماس‌های افراد بوده است. نتایج شبیه‌سازی‌شده توسط داده‌های تلفن همراه بر مبنای فواصل زمانی کوتاه، به داده‌های واقعی استخراج شده حاصل از حلقه‌های القایی نزدیک بوده است. به همین منظور محققین این پژوهش، استفاده از داده‌های ارزان‌قیمت تلفن همراه را بجای سایر ابزارهای شمارشی، نظیر دوربین و حلقه القایی، محتمل و امکان‌پذیر دانسته‌اند. مطالعه گنزالس و همکاران [González et al., 2008] نیز از اولین مطالعات انجام‌شده در زمینه استفاده از داده‌های تلفن همراه در ابعاد کلان می‌باشد. پایگاه داده این مطالعه شامل ۱۰۰ هزار کاربر بوده که اطلاعات پیامک‌های ایشان نیز علاوه بر تماس‌ها موجود است. گنزالس در این پژوهش با استفاده از یک تابع نمایی شعاع ژیراسیون حرکتی افراد را مدل‌سازی کرد. سپس افراد را بر اساس شعاع ژیراسیون حرکتی محاسبه‌شده دسته‌بندی نمود تا شباهت حرکت انسان‌ها را بررسی کند. در انتها گنزالس و همکاران مدل‌سازی رفتار افراد را با استفاده از داده‌های تلفن همراه امکان‌پذیر دانسته و استفاده از آن‌ها را به عنوان یکی از ابزارهای ارتقاء شبکه حمل‌ونقل شهری نام برده‌اند. پس از گنزالس، کالابرس و همکاران [Calabrese, Di Lorenzo, et al., 2011] در شهر بوستون مطالعاتی را درباره چگونگی شناسایی نقاط فعالیت افراد و دسته بندی نوع فعالیت آن‌ها با استفاده از ۱۳۰ میلیون رکورد و داده‌های تلفن همراه مثلث‌سازی‌شده به همراه یک میلیون شناسه متمایز، انجام دادند. در این مطالعه رکوردهای ثبت شده بر اساس تعاملات کاربر با شبکه تلفن همراه شامل تماس، پیامک، و اتصال به اینترنت بوده است. در این مطالعه به منظور بررسی صحت نتایج، از مقایسه طول سفر، جمعیت برآورد شده و نرخ سفر روزانه، استفاده شده است. کالابرس در همین مطالعه که به شناسایی نقاط فعالیت افراد پرداخت، امکان سنجی استخراج ماتریس مبدا–مقصد از داده‌های تلفن همراه را نیز بررسی کرده است. در این مطالعه، صحت‌سنجی داده‌های تلفن همراه از طریق بررسی فاصله زمانی میان تماس‌ها، بررسی نرخ سفر پایگاه داده برای روزهای تعطیل و غیرتعطیل، مقایسه طول سفرهای استخراج‌شده و مطالعات طرح جامع و مقایسه جمعیت مقیاس‌شده افراد ساکن شناسایی‌شده توسط داده‌های تلفن همراه و مطالعات طرح جامع ، انجام گرفته است. تا به اینجا محققین بیشتر به دنبال اثبات کارآیی این نوع داده‌ها بودند. در برخی مطالعات به ارائه و توسعه روش‌های استخراج سفرها و مقابله با مشکلات موجود در این زمینه پرداخته‌اند. به عنوان مثال اقبال و همکاران [Iqbal et al., 2014] برای ساخت ماتریس مبدا–مقصد با استفاده از داده‌های تلفن همراه، از رویکردی مبتنی بر شبیه‌سازی استفاده کردند. این مطالعه در محدوده شهر داکا با استفاده از پایگاه داده متشکل از ۶,۹ میلیون کاربر انجام شده است. در این مطالعه، به منظور شمارش داده‌های واقعی حجم، از ویدئوهای ترافیکی ثبت‌شده توسط دوربین‌های نظارتی استفاده شده است. در این پژوهش ابتدا ماتریس گذرای مبدا–مقصد[۷] که توسط وانگ ابداع شده است [Jing et al., 2011]، تشکیل شده است. این

ماتریس که بر مبنای سفر بین آنتن‌های مخابراتی[8] می‌باشد، به ماتریس گره‌به‌گره[9] (بین نقاطی که داده‌های دوربین در دسترس است)، تبدیل شده است. به منظور تخصیص مبدا و مقصد هر یک از سفرها، که مختصات و شناسه یک آنتن مخابراتی می‌باشد، بر اساس نزدیکی فاصله آنتن‌ها به گره‌ها، عمل شده است. همچنین برای تعمیم ماتریس گذرا به مقدار واقعی، از یک رویکرد شبیه‌سازی و بهینه‌سازی مبتنی بر تفاوت حجم واقعی و شبیه‌سازی، استفاده شده است.

با گذشت زمان محققین به بهبود کیفیت داده‌ها که نتیجه مستقیمی بر نتایج دارند، علاقه‌مند شدند. در همین راستا الکساندر و همکاران [Alexander et al., 2015] با استفاده از داده‌های تلفن همراه مثلث‌سازی‌شده به بررسی ماتریس‌های مبدا–مقصد بر اساس هدف سفر پرداختند. پایگاه داده‌های تلفن همراه استفاده‌شده در این مطالعه شامل ۲ میلیون کاربر در مدت ۲ ماه بوده است. ابتدا با پردازش داده‌های خام تلفن همراه، نقاط توقف و عبوری برای هر کاربر استخراج شده است. نحوه استفاده از این الگوریتم بر اساس قیود مکانی و زمانی بوده؛ به این صورت که ابتدا نقاطی که در فاصله‌ای کمتر از یک فاصله مکانی مشخص باشند به عنوان یک خوشه اولیه شناسایی می‌شوند. سپس مدوید[10] این خوشه اولیه به عنوان نماینده خوشه استفاده می‌شود. مدوید مفهومی مشابه میانگین یا مرکز سطح است، با این تفاوت که حتماً باید یکی از اعضای خود خوشه باشد. در ادامه بدون توجه به ترتیب زمانی، این مدوید به خوشه ثانویه تبدیل شده و مجددا مدوید این خوشه‌های ثانویه به عنوان نماینده نهایی نقاط انتخاب می‌شوند. بعد از شناسایی و پاک‌سازی پایگاه داده از افراد با میزان فعالیت کم سفرها، ماتریس‌ها بر اساس سه هدف خانه–محل کار، خانه–سایر و هیچ‌سرخانه، محاسبه شده‌اند. در ساخت ماتریس‌ها از مفهومی به نام روز موثر، استفاده شده است. این مفهوم بیانگر این است که، امکان دارد سفرهای افراد گاهاً به ساعاتی از روز بعد منتقل شوند به همین منظور هر روز، به صورت موثر، از ساعت ۳ صبح تا ساعت ۳ صبح روز بعد تعریف شده است. همچنین فرض شده است که ابتدا و انتهای سفرهای هر شخص از خانه شروع، و به خانه ختم شده باشند. بعد از محاسبه ماتریس سفر برای هر شخص، به منظور تعمیم سفرهای فرد به کل جمعیت ناحیه ساکن، از ضریب تعمیم ماتریس استفاده شده است. در این مطالعه از شاخص‌هایی نظیر نرخ سفر و مقایسه آن با مطالعات انجام‌شده در محدوده بوستون، مقایسه زمان شروع سفرها، طول سفرهای انجام‌شده، جمعیت تعمیم‌داده‌شده هر ناحیه و حجم سفرهای تخصیص‌داده‌شده به هر لینک، برای ارزیابی استفاده شده است. نتیجه مهم حاصل از این مطالعه، تاثیر ابعاد ناحیه‌بندی در تطابق سفرهاست. هر چه سفرها در نواحی بزرگتری مورد مطالعه قرار گیرند، نتایج به مطالعات جامع انجام‌شده، نزدیکتر خواهند بود. جولاک و همکاران [Çolak et al., 2015] نیز به منظور سنجش کارایی داده‌های تلفن همراه، به بررسی امکان ساخت ماتریس‌های مبدا–مقصد، با استفاده از داده‌های دو شهر ریو و بندر سانفراسیسکو، پرداختند. پایگاه داده استفاده‌شده به منظور بررسی بندر سانفرانسیسکو، شامل ۸ میلیارد رکورد به همراه مختصات مثلث‌سازی‌شده و ۲ میلیون کاربر متمایز، و پایگاه داده استفاده‌شده برای شهر ریو نیز، شامل ۱ میلیارد رکورد با مختصات آنتن‌مبنا به همراه ۲٫۸ میلیون کاربر متمایز می‌باشد. در این مطالعه به منظور استخراج نقاط توقف از یک الگوریتم مبتنی بر فواصل زمانی و مکانی استفاده شده است. همچنین به منظور اصلاح زمان ورود و خروج، از روش احتمالی برای مطالعه سانفرانسیسکو، و از روش‌های ساده مبتنی بر میانگین وزنی برای شهر ریو، استفاده شده است. در این مطالعه با استفاده از مقایسه جمعیت تعمیم‌داده‌شده با جمعیت واقعی در هر ناحیه آماری، طول سفر، الگوی تخصیص حجم در معابر مختلف، و بررسی زمان ورود و خروج به نقاط فعالیت، بر حسب نوع آن، ارزیابی انجام شده است.

بونل و همکاران [Bonnel et al., 2018] با هدف ساخت ماتریس مبدا–مقصد، مطالعه‌ای را به مدت یک روز بر روی

پایگاه داده‌ای با ۱,٦۲ میلیون کاربر انجام دادند. به منظور بررسی چگونگی انجام سفرها، ابتدا به وسیله یک روش با قید زمانی، نقاط توقف شناسایی شدند. بعد از شناسایی سفرها و استخراج ماتریس‌ها با استفاده از ضرایب توسعه میانگین کل جمعیت منطقه مورد مطالعه، ماتریس‌ها به کل جمعیت تعمیم داده شدند. در ادامه این مطالعه، در سال ۲۰۲۰، بونل با استفاده از رویکردی مشابه، به بررسی مطالعات ماتریس مبدا–مقصد پرداخت؛ با این تفاوت که در این مطالعه به منظور حذف نقاط با فعالیت زیاد، از رویکردی مبتنی بر آنتروپی در بازه‌های زمانی یک ساعته استفاده شده است. مقایسه سفرهای بدست‌آمده بین نواحی مختلف با مطالعات حمل‌ونقلی انجام‌شده در گذشته نشان داد که استفاده از داده‌های تلفن همراه، به عنوان ابزار ارزیابی و ساخت ماتریس های مبدا–مقصد، امکان‌پذیر می‌باشد [Fekih et al., 2020]. امکان بررسی تغییرات الگوهای حرکتی ناشی از سوانح خاص از دیگر مزایای استفاده از این نوع داده‌ها است که اهمیت آن پس از همه‌گیری ویروس کرونا (COVID-19) بیش از پیش قابل درک است، چرا که این داده‌ها در مدت زمان کم و با هزینه پایین قابل جمع‌آوری هستند. سینتیا و همکاران [Cintia et al., 2020] با استفاده از این داده‌ها به بررسی وضعیت قبل و بعد از همه‌گیری ویروس کرونا در ایتالیا و تاثیرگذاری محدودیت‌های کرونایی پرداختند. یکی از نتایج مهم این مطالعه برای مثال آن است که تغییرات قابل ملاحظه در الگوهای جابجایی ناشی از اعمال محدودیت های کرونایی حدوداً یک هفته پس از اجرای این محدودیت‌ها اتفاق می‌افتند. این فاصله زمانی می‌تواند در تصمیم گیری در خصوص اقدامات مقابله‌ای برای کاهش همه‌گیری حائز اهمیت باشد.

اخیراً محققان دانشگاه توکیو [Imai et al., 2021] به کمک داده‌های عملیاتی شبکه‌های مخابراتی به استخراج ماتریس‌های مبدأ–مقصد در سطح ملّی پرداختند. داده‌های مورد استفاده در این مطالعه، شامل داده‌های ۷٦ میلیون کاربر برای بازه زمانی یک‌روزه بوده است. همچنین این داده‌ها از نوع داده‌های عملیاتی شبکه تلفن همراه بوده و هر یک ساعت، موقعیت کاربر در پایگاه داده مورداستفاده در این مطالعه، به‌روز می‌شده است. ازجمله مزایای داده‌های مورد استفاده در این مطالعه، گنجاندن اطلاعات جنسیتی، سنی و محل سکونت افراد در پایگاه داده‌های موبایل بوده است. بدین ترتیب امکان استفاده از ضرایب تعمیم دقیق‌تر و شناسایی دقیق‌تر محل زندگی افراد، میسر شده است. به منظور صحت‌سنجی نتایج، از داده‌های مربوط به پرسشگری سفر در این کشور استفاده شده است. در این مطالعه از یک روش سه‌مرحله‌ای برای استخراج اطلاعات سفر استفاده شده است. در اولین مرحله حذف اطلاعات هویتی افراد به منظور حفظ حریم شخصی، در گام بعد استخراج ماتریس مبدأ–مقصد در سطح مناطق، و در نهایت حذف سلول‌هایی از ماتریس که دارای اطلاعات کافی برای تخمین ماتریس نبوده و احتمال خطا در آن‌ها وجود دارد، صورت گرفته است. همچنین برای اعتبارسنجی نتایج، از سه معیار توزیع طول سفرها، نرخ انجام سفر، و مقایسه ماتریس های مبدا–مقصد به دست آمده از مطالعات پرسشگری سفر، استفاده شده است.

نکته مهم دیگری که در استفاده از داده‌های CDR مورد توجه محققین بوده است، وجود یک منبع قابل اعتماد برای اعتبارسنجی نتایج حاصله از این نوع داده‌هاست. پورمرادناصری و همکاران [Pourmoradnasseri et al., 2019] در مطالعه‌ای با استفاده از داده‌های CDR و ثبت موقعیت بازدیدکنندگان (VLR) روشی را بر مبنای زنجیره مارکوف به منظور استخراج الگوهای حرکتی افراد ارائه کردند. آن‌ها اعلام کردند که اگر منبع قابل‌اعتمادی که نتایج نزدیک‌تر به نتایج CDR را ارائه دهد وجود داشته باشد، نتایج بهتری ارائه خواهند شد. در این زمینه لندمارک و همکاران [Dypvik Landmark et al., 2021] مطالعه‌ای را برای پیدا کردن بهترین منبع برای اعتبارسنجی نتایج حاصل از CDR انجام دادند. نهایتاً ایشان نشان دادند که خروجی‌های CDR همبستگی بالایی با خروجی‌های مطالعات جامع دارد. اما این همبستگی صرفاً در نمایش الگوی سفرها

عملکرد خوبی داشته و تعداد سفرها را درست پیشبینی نمی‌کند. آن‌ها همچنان نتیجه گرفتند که هرکجا الگوریتم‌هایی که برای رعایت حریم شخصی افراد استفاده می‌شوند سختگیرانه‌تر باشند، استفاده از داده‌های CDR به عنوان جایگزین داده‌های سنتی مناسب نخواهد بود و بهتر است به عنوان داده‌های کمکی مورد استفاده قرار گیرند.

در داخل کشور نیز تاکنون مطالعات حمل‌ونقلی پیرامون داده‌های شبکه تلفن همراه انجام شده است، اما این مطالعات پیرامون داده‌های PLU بوده و همچنین مدت زمان تحلیل داده‌ها عموماً کوتاه، به مدت حداکثر سه روز بوده است. علت این موضوع نیز ماهیت داده‌های PLU است که مانند داده‌های تلفن همراه، توسط اپراتورها جمع‌آوری نمی‌شوند و صرفاً برای مقاصد پژوهشی انجام می‌گیرد. به همین علت، استفاده از داده‌های PLU، هزینه بیشتری نسبت به داده‌های تلفن همراه دارد. به عنوان مثال، در سال ۹۵، زاهدی با استفاده از داده‌های ۳۰۰ هزار کاربر در شهر شیراز، که این داده‌ها با فاصله زمانی ۱۰ دقیقه توسط اپراتور همراه اول ثبت شده بودند، به بررسی الگوی فعالیتی افراد پرداخت. روش استفاده‌شده برای شناسایی نقاط فعالیت افراد با الهام‌گیری از رفع پدیده پینگ‌پنگی[11] داده‌های شبکه بی‌سیم بوده است. پدیده پینگ‌پنگی به جابجایی متواتر نقطه اتصال مورد استفاده یک تلفن همراه گفته می‌شود که می‌تواند تحلیل را با شناسایی سفرهای غیرواقعی، با خطای قابل‌توجهی روبرو سازد. همچنین در این مطالعه به دلیل دسترسی به اطلاعات کاربری زمین، نوع محل فعالیت با توجه به سهم هر کاربری تعیین شده است [Zahedi et al. 1395]. در مطالعه‌ای دیگر در سال ۹۷، افتخار با استفاده از داده‌های مشابه، به بررسی ماتریس مبدا–مقصد برای ساعات اوج پرداخت [Eftekhar et al. 1397]. همچنین در سال ۹۷، بیگدلی با استفاده از داده‌های مشابه، به بررسی تطبیق داده‌های استخراج شده از داده‌های ثبت‌شده شبکه تلفن همراه با داده‌های کمان‌های شمارش حجم و اصلاح نتایج آن پرداخت [Bigdeli et al. 1397]. علاوه بر این، در سال ۹۸، نوری با استفاده از نتایج مطالعات قبلی به بررسی نقشینه‌های استخراج شده در سطح شهر شیراز پرداخت [Noori et al. 1395]. در تمامی مطالعات انجام شده، روش‌های شناسایی استفاده‌شده برای داده‌های PLU را، مناسب پردازش داده‌های تلفن همراه نمی‌دانستند. البته کارهای دیگری با استفاده از انواع داده‌های دیگر مثل GPS نیز صورت گرفته است. شکری و عباسپور در سال ۹۵ با استفاده از داده‌های GPS به استخراج اطلاعات ترافیکی در شهر تهران پرداختند [Shokri & Abbaspour, 2017]. نرخ نمونه‌برداری در این مطالعه نیز پایین بوده است.

نهایتاً هدف از انجام این مطالعه، بررسی کارایی این داده‌ها، که تفکیک‌پذیری مکانی کمی دارند، در مطالعات حمل‌ونقلی است. به همین منظور، در این پژوهش به امکان‌سنجی بررسی و شناسایی رفتار سفر افراد در سطح شهر تهران، با استفاده از برآورد ماتریس مبدا–مقصد پرداخته شده است. داده‌های مورد استفاده در این مطالعه به صورت خام، آماده برای استفاده در تحلیل‌های حمل‌ونقلی نبوده و به همین علت ابتدا لازم است با استفاده از روش‌های نوین کار با کلان‌داده‌ها، داده‌ها آماده تحلیل شوند. سپس الگوهای جابجایی، ماتریس‌های مبدأ–مقصد و تولید–جذب، و نرخ سفرها، برای سفرهای خانه‌مبنای کاری، غیرکاری، و سایر اهداف سفر استخراج شده‌اند. به منظور اعتبارسنجی، نرخ سفر و تعداد سفرهای بدست‌آمده از تحلیل داده‌های تلفن همراه شهر تهران در بازه یک‌ماهه با نتایج حاصل از مطالعات جامع حمل و نقل شهر تهران مقایسه گردیده است. ماتریس‌های سفر نیز با توجه به تغییرات در کاربری‌ها و الگوهای رفتاری موجود، کنترل شده‌اند.

۲. بررسی پایگاه داده مورد استفاده

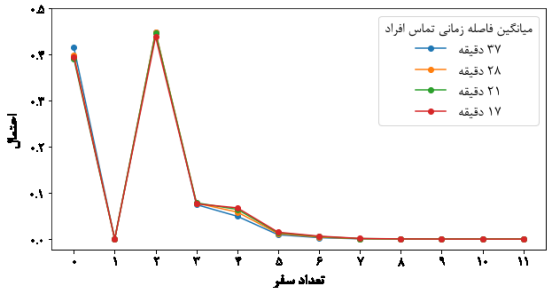
داده‌های تلفن همراه که در این مطالعه مورداستفاده قرار گرفته‌اند، توسط اپراتورهای تلفن همراه جمع‌آوری می‌شوند. این داده‌ها به صورت خام، کاربرد حمل و نقلی ندارند و لازم است برای قابل استفاده شدن، پیش‌پردازش بر روی داده‌ها صورت گیرد.

داده‌های تلفن همراه استفاده‌شده در مطالعات پیشین با توجه به هدف مطالعه، که می‌تواند بررسی حرکت افراد در بازه‌های کوتاه‌مدت یا بلندمدت باشد، از بازه‌های یک‌روزه [Bonnel et al., 2018; Imai et al., 2021] تا یک‌ساله [Barboza et al., 2021] جمع‌آوری شده‌اند. داده موجود در مطالعه پیش‌رو مربوط به یک بازه یک‌ماهه از تابستان سال ۱۳۹۸ می‌باشد. این بازه یک‌ماهه امکان شناسایی محل اصلی فعالیتی افراد (و به تبع آن تفکیک اهداف سفر) و همچنین مطالعه رفتار سفر به تفکیک روزهای کاری و غیرکاری را فراهم می‌سازد. پایگاه داده مذکور شامل اطلاعات ۱۱،۹۲۸،۱۲۲ نفر در قالب ۱،٦٥١،۳۷٦،۲۳٤ رکورد اطلاعاتی مربوط به بزرگترین اپراتور تلفن همراه کشور (همراه اول)، است. هر رکورد در پایگاه داده شامل تاریخ رویداد (تماس‌ها و پیامک‌ها)، شناسه متمایز هر کاربر، زمان رویداد، و طول و عرض جغرافیایی آنتن سرویس‌دهنده به آن است. نوع داده‌ها در این مطالعه آنتن‌مبنا بوده که با بررسی دقیق‌تر مشخص گردید که طول و عرض جغرافیایی ثبت‌شده برای هر آنتن در واقع حاصل تجمیع چند آنتن در یک نقطه می‌باشد. هرچند که این مسئله از تفکیک‌پذیری مکانی داده‌ها می‌کاهد و خروجی نهایی را تحت تاثیر قرار می‌دهد، اما نگرانی‌ها در مسئله حفظ حریم شخصی افراد را کاهش می‌دهد. این مسئله در مطالعه اخیر جیانگ و همکاران [W. Jiang & Luo, 2022] نیز مورد اشاره قرار گرفته است که تجمیع موقعیت آنتن‌ها در سطوح وسیع‌تر و عدم نمایش مختصات دقیق می‌تواند در حفظ حریم شخصی کاربران موثر باشد.

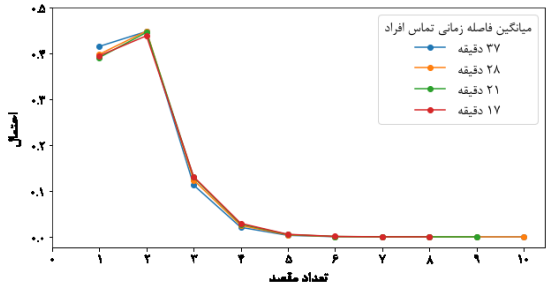
عامل مهم دیگری که بر کیفیت مطالعه اثرگذار است، تعداد و فواصل (تفکیک‌پذیری مکانی) آنتن‌ها است. در پایگاه داده موجود، ۷۱۲ آنتن با میانگین فاصله ٤٨٦ متر وجود دارند. برای بررسی بهتر پراکندگی آنتن‌ها و ناحیه تحت پوشش آن‌ها، از چندضلعی ورونوی[۱۲] استفاده شده است. چندضلعی ورونوی یک محدوده فرضی اطراف آنتن‌های مخابراتی، تحت عنوان ناحیه تحت پوشش هر آنتن، ترسیم می‌کند. نواحی مرکزی و شمالی شهر تهران، تراکم آنتن بالاتری نسبت به سایر مناطق دارند. علت این موضوع را نیز می‌توان تمایل اپراتورهای مخابراتی به کاهش بار هر آنتن در مناطق مرکزی و شمالی که تراکم جمعیت بالایی دارند، دانست. همچنین باید در نظر داشت که محدوده تحت پوشش واقعی هر آنتن، با توجه به تجمیع موقعیت آنتن‌ها، با احتمال زیادی با منحنی‌های ترسیم شده تفاوت دارد.

نکته دیگری که علاوه بر تفکیک‌پذیری مکانی باید مورد توجه قرار گیرد، تفکیک‌پذیری زمانی داده‌ها است. افرادی که فاصله زمانی تماس‌هایشان از یک آستانه مشخصی بیشتر باشد، الگوی جابجایی قابل‌اتکایی ندارند، چرا که ممکن است برخی سفرهایشان در بازه‌ای که رکوردی در پایگاه داده ندارند، صورت گرفته باشد [Board et al., 2018]. در این مطالعه با بررسی داده‌ها، افرادی که فاصله بین تماس آنها کمتر از یک ساعت بوده است به عنوان داده‌های قابل اطمینان، جدا شدند. فاصله یک‌ساعته بین دو رکورد، با توجه به مطالعات پیشین صورت گرفته در این زمینه انتخاب شده است [Board et al., 2018; Bonnel et al., 2018; Fekih et al., 2020]. تعداد این افراد ٤۰۰،۳۷۷ نفر است. میانگین بازه زمانی بین دو تماس متوالی برای این افراد، حدود ۳۸ دقیقه می‌باشد که برای استخراج سفرها، مورد قبول است [Board et al., 2018]. با این جداسازی، بیم آن می‌رود که رفتار سفر افرادی که جدا شده‌اند، به طور معناداری با افرادی که باقی مانده‌اند، تفاوت داشته باشد. در این مورد، ژیانگ و همکاران در مطالعه‌ای نشان دادند که رابطه معناداری بین میزان استفاده از تلفن همراه و الگوهای جابجایی و تعداد سفر افراد وجود ندارد [S. Jiang et al., 2017]. جهت بررسی صحت این مدعا، فراوانی تعداد سفرهای روزانه در گروه‌های مختلف کاربران در شکل (۱)، فراوانی تعداد مقاصد متفاوت در سفرهای روزانه افراد در گروه‌های مختلف کاربران در شکل (۲)، و نمودار توزیع فراوانی طول سفرهای

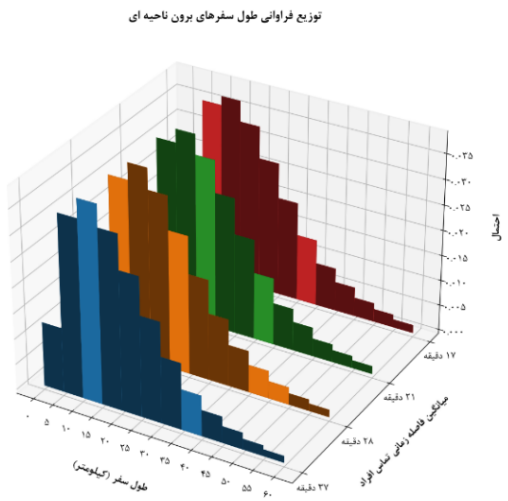
روزانه بین گروه‌های مختلف کاربران برون‌ناحیه‌ای در شکل (۳)، و درون‌ناحیه‌ای در شکل (٤) مورد مقایسه قرار گرفتند.



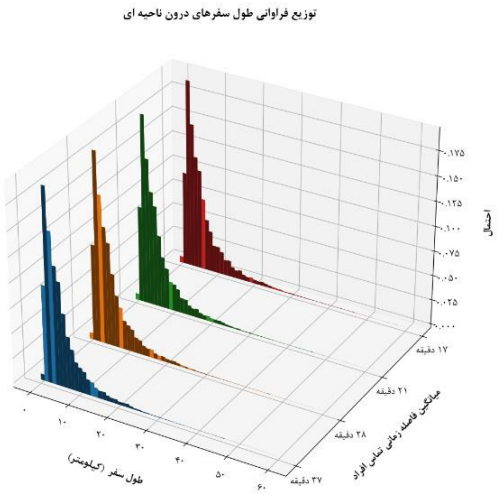
شکل ۱. مقایسه فراوانی تعداد سفرهای روزانه در گروه‌های مختلف کاربران



شکل ۲. مقایسه فراوانی تعداد مقاصد متمایز روزانه در گروه‌های مختلف کاربران



شکل ۳. مقایسه نمودار توزیع فراوانی طول سفرهای روزانه بین گروه‌های مختلف کاربران برون‌ناحیه‌ای



شکل ٤. مقایسه نمودار توزیع فراوانی طول سفرهای روزانه بین گروه‌های مختلف کاربران درون‌ناحیه‌ای

همانطور که مشاهده می‌شود، این نمودارها علاوه بر نمایش عدم وابستگی الگوی سفر افراد به فاصله زمانی تماس‌ها، حاوی اطلاعات دیگری نیز هستند که به نظر منطقی می‌رسند. اکثر افراد ۱ یا ۲ مقصد متمایز داشته‌اند، که به ترتیب معادل ۰ یا ۲ سفر در روز است. افراد اغلب یا سفر نکرده‌اند، و یا یک سفر کاری انجام داده و به منزل بازگشته‌اند. طول سفر افراد نیز مستقل از فاصله تماس آن‌ها بوده است. این موارد می‌توانند نمایانگر منطقی بودن داده‌های ورودی باشند.

۳. روش‌شناسی پژوهش

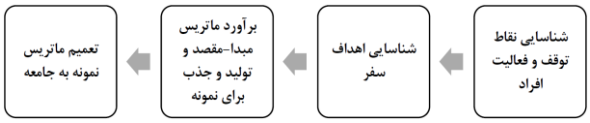
با توجه به اینکه داده‌های موجود حجم بالایی داشته و داده‌های پرت و اشتباه در آن زیاد است، ابتدا با استفاده از روش‌های موجود مدیریت کلان‌داده‌ها، این داده‌ها حذف شده، و نقاط توقف و فعالیت افراد به منظور شناسایی مبادی و مقاصد ایشان، مورد بررسی قرار گرفتند. با توجه به آنکه داده‌های تلفن همراه مختص تحلیل‌های حمل و نقلی جمع‌آوری نمی‌شوند و همچنین با توجه به حجم بسیار بالای داده و وجود خطاهایی در ذخیره‌سازی، عملیات پیش‌پردازش به منظور حذف داده‌های پرت و مشکل‌دار انجام شد. نهایتاً داده‌ها با مشخصات ذکرشده در جدول (۱) مورد استفاده قرار گرفتند. اصلاحات لازم بر روی

ساختار داده‌ها و همچنین تحلیل آن‌ها، به منظور حذف داده‌های پرت، ایرادات موجود در داده‌ها، و غیره، صورت گرفته است. به عنوان مثال ایراداتی که در اکثر داده‌های کلان مشاهده می‌شود، از قبیل سلول‌های خالی، کاراکترهای بی‌ربط، جابجایی در برخی سلول‌ها و غیره، در داده‌های ما نیز مشهود بوده است. برخی مشکلات دیگر نیز وجود داشتند که زمان پردازش داده‌ها را با تاخیرات طولانی همراه می‌ساختند. به عنوان مثال، برخی سلول‌ها در بخش موقعیت کاربران دارای تعداد ارقام نامتناهی اعشاری بودند. این در حالیست که موقعیت‌ها در اکثر سلول‌ها دارای سه رقم اعشار بودند. این مشکل احتمالاً در فرآیند تجمیع داده‌ها رخ داده است. با استفاده از فنون کار با کلان‌داده‌ها و بسته‌های نرم‌افزاری متن‌باز، این مشکلات مرتفع شدند. تو ضیح بیشتر درباره پالایش داده‌ها در این مقال نمی‌گنجد.

پژوهش پیش‌رو در چهار گام انجام شده است. در گام اول، نقاط توقف و فعالیت افراد شناسایی شده و در گام دوم افراد بر اساس اهداف سفر دسته‌بندی شده‌اند. در گام سوم و چهارم به برآورد ماتریس مبدا–مقصد برای نمونه داده‌های دردست و تعمیم آن به جامعه که تمام افراد شهر تهران هستند، پرداخته شده است. در شکل (۵) گام‌های روش پژوهش به صورت کلی مشخص شده‌اند که در ادامه به جزئیات هرکدام پرداخته شده است.

جدول ۱. اطلاعات آماری مربوط به محدوده مورد مطالعه و تفکیک‌پذیری زمانی و مکانی داده‌ها

آماره	مشخصات
تعداد آنتن‌های مخابراتی	۷۱۲ عدد
میانگین فاصله آنتن‌ها	٤٨٦ متر
میانگین فاصله زمانی بین تماس‌های افراد بعد از حذف داده‌های غیرقابل استفاده	۳۸ دقیقه
تعداد کاربران قبل از حذف داده‌های غیرقابل استفاده	۱۱,۱۹۸,۱۲۲نفر
تعداد کاربران بعد از حذف داده‌های غیرقابل استفاده	٤۰۰,۳۷۷ نفر
تعداد ساکنین در شهر تهران در نمونه اصلاح‌شده	۳٤٦,۸۳۲ نفر
تعداد رکوردها قبل از حذف داده‌های غیرقابل استفاده	۱,٦٥۱,۳۷٦,۲۳٤
تعداد رکوردها بعد از حذف داده‌های غیرقابل استفاده	٥۱۱,٦۹٥,۰۰۰
تعداد رکوردهای ساکنین	٤۲۸,۹۳۰,۷۹۳



شکل ۵. گام‌های روش پژوهش

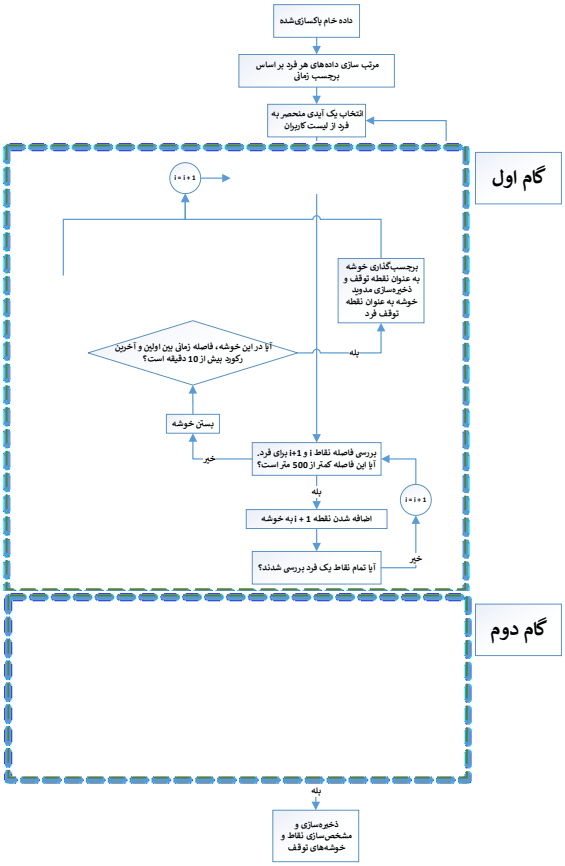
۱-۳ شناسایی نقاط توقف و فعالیت افراد

تشخیص نقاطی که افراد در آن توقف و یا فعالیت داشته‌اند، اولین گام در شناسایی الگوی سفر افراد است. همانطور که بیان شد در پایگاه داده موجود، موقعیت دقیق هر فرد مشخص نیست و صرفاً موقعیت آنتنی که فرد به آن متصل شده، ثبت شده است. بنابراین تجمیع داده‌های خام در سطح نواحی، با خطا همراه خواهد بود. به همین دلیل نقاط توقف و فعالیتی با استفاده از الگوریتم‌هایی که در ادامه توضیح داده می‌شوند، شناسایی گردیده و پس از تعیین محل حدودی توقف و فعالیت افراد، این نقاط به مناطق نسبت داده شدند. در این مطالعه از یک الگوریتم دومرحله‌ای جهت شناسایی نقاط توقف[۱۳] و فعالیت افراد[۱٤] استفاده شده است (شکل (٦)). بدین منظور، از روش شبکه‌بندی و الگوریتم‌های یادگیری ماشین استفاده شده است. در گام اول پس از مرتب کردن رکوردهای هر فرد بر مبنای توالی زمانی، با استفاده از روشی که تشریح خواهد شد به کنترل شرط مکانی و زمانی برای هر فرد می‌پردازیم. در گام دوم با استفاده از یک الگوریتم خوشه‌بندی بدون توجه به توالی زمانی از نظر مکانی خوشه‌ها را کنترل می‌کنیم. در ادامه به تشریح این روش پرداخته می‌شود. فرض می‌شود که مجموعه داده‌های مشاهده‌شده برای یک فرد در روز اول، به شکل یک توالی زمانی باشد (رابطه (۱)).

$$D_i = (d_i(1), d_i(2), d_i(3), \ldots, d_i(n)) \quad (۱)$$

در این توالی، هر یک از مولفه‌ها خود شامل زمان اتصال فرد به آنتن مخابراتی ($t_{(k)}$)، طول جغرافیایی آنتن ($x_{(k)}$)، و عرض جغرافیایی آنتن ($y_{(k)}$)، می‌باشند (رابطه (۲)).

$$d_i(k) = (t_{(k)}, x_{(k)}, y_{(k)}) \quad (۲)$$



شکل ٦. فلوچارت شناسایی نقاط توقف افراد

در مرحله اول از این الگوریتم، نقطه اول ($d_i(1)$) از نظر مکانی با نقطه دوم ($d_i(2)$) مقایسه می‌گردد. اگر فاصله دو نقطه کمتر از ۵۰۰ متر باشد، این دو نقطه، یک خوشه را تشکیل می‌دهند. در گام بعد، نقطه بعدی فرد ($d_i(3)$) با نقطه ابتدای خوشه مقایسه شده و اگر همچنان فاصله این دو نقطه کمتر از ۵۰۰ متر باشد، نقطه جدید به خوشه جاری اضافه می شود. این کنترل تا جایی صورت می‌گیرد که نقطه جدید فاصله‌ای بیش از ۵۰۰ متر با نقطه اول خوشه داشته باشد. در این صورت، نقطه جدید خود نقطه ابتدایی خوشه جدید شده و مراحل پیشین برای خوشه جدید برای نقاط بعدی فرد تکرار می‌شود. این عمل برای تمام نقاط فرد در پایگاه داده با توالی که ذکر شد صورت می‌گیرد. فاصله ۵۰۰ متر در این مطالعه با توجه به میانگین فاصله آنتن‌ها که ٤٨٦ متر است، برای غلبه بر پدیده پینگ‌پنگی مطابق شکل (۷–الف)، در نظر گرفته شده است. برای محاسبه فاصله بین نقاط نیز از رابطه هاورساین[۱۵] به شرح رابطه (۳)، استفاده شده است.

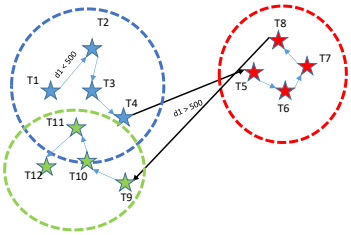
$$d_{ij}=2R\arcsin\left(\sqrt{\sin^2\left(\frac{Lat(j)-Lat(i)}{2}\right)+\cos(Lat(i)).\cos(Lat(j)).\sin^2\left(\frac{Lng(j)-Lng(i)}{2}\right)}\right) \quad (۳)$$

در این رابطه، با استفاده از طول جغرافیایی[۱۶]، عرض جغرافیایی[۱۷]، و شعاع زمین (R) که ٦٧٣١ کیلومتر لحاظ شده است، فاصله دو نقطه محاسبه می‌گردد.

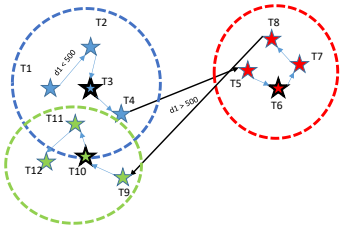
در گام بعد، مدوید هر خوشه به عنوان نماینده آن خوشه در نظر گرفته می‌شود (شکل (۷–ب)). علت استفاده از مدوید نیز، حذف نوسانات احتمالی ناشی از تبادلات بین آنتن‌های مخابراتی[۱۸] می‌باشد. پس از تشکیل هر خوشه، فاصله زمانی بین نقطه اول و آخر هر خوشه محاسبه می‌گردد. اگر این فاصله زمانی از ۱۰ دقیقه بیش‌تر باشد، موقعیت مدوید آن خوشه به عنوان یک نقطه توقف در نظر گرفته می‌شود و در غیر این صورت، آن خوشه به عنوان مجموعه‌ای از نقاط عبوری در نظر گرفته می‌شود (شکل (۷–ج)). این شرط زمانی، ضمن جلوگیری از وقوع پدیده پینگ‌پنگی، از شناسایی جابه‌جایی‌های افراد در محل فعالیت به عنوان سفر نیز جلوگیری می‌کند.

استفاده از نقاط توقف شناسایی‌شده در مرحله قبل دارای یک اشکال جزیی خواهد بود. برای مثال، با توجه به اینکه ممکن است یک فرد در روزهای مختلف در یک نقطه فعالیتی منحصربه‌فرد و یکتا (مانند منزل) به آنتن‌های مختلفی متصل شود، این احتمال وجود دارد که در روزهای مختلف، چند موقعیت متفاوت برای خانه فرد ثبت شود. به منظور رفع این خطا، پس از شناسایی همه نقاط توقف برای یک فرد، با استفاده از یک الگوریتم خوشه‌بندی مکانی، نقاط توقف نزدیک به هم برای آن فرد تجمیع شده و نقاط فعالیتی وی را تشکیل می‌دهند (شکل (۷–د)). برای تجمیع نقاط توقف از الگوریتم خوشه‌بندی سلسله‌مراتبی تجمعی[۱۹] استفاده شده است. مزیت این الگوریتم

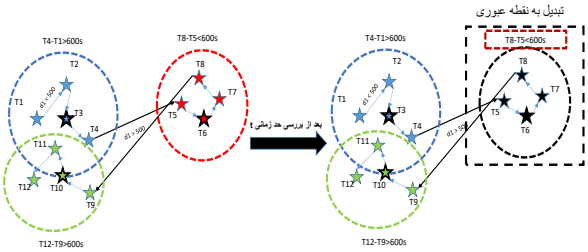
عدم نیاز به تعیین تعداد خوشه‌های موجود است، که با توجه به تفاوت تعداد نقاط توقف برای افراد مختلف، و یا افراد یکسان در روزهای مختلف، در این مطالعه بکار گرفته شده است.



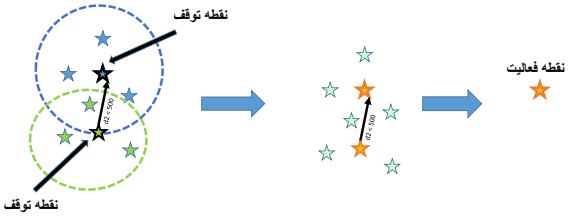
(الف): خوشه‌بندی نقاط مستعد محل فعالیت با اعمال شرط مکانی



(ب): انتخاب مدوید هر خوشه مستعد محل فعالیت به عنوان نماینده آن خوشه



(ج): بررسی شرط زمانی خوشه های مستعد محل فعالیت



(د): اعمال الگوریتم HAC و شناسایی محل فعالیت افراد

شکل ۷. فرآیند شناسایی نقاط توقف و فعالیت افراد

الگوریتم خوشه‌بندی سلسله مراتبی، برای لحاظ کردن فواصل، شاخص‌های[20] مختلفی دارد که در این مطالعه، از شاخص فاصله اقلیدسی به شرح رابطه (٤) استفاده شده است.

$$\| a-b \|_2 = \sqrt{\sum_i (a_i - b_i)^2} \qquad (٤)$$

در این رابطه، a و b، نقاط مورد بررسی هستند. الگوریتم مورد استفاده برای محاسبه فاصله بین نقاط، نیاز به پیوند[21] دارد .در انتخاب پیوند الگوریتم، از پیوند میانگین غیروزنی استفاده شده است. پیوند میانگین غیروزنی، فاصله نقاط یک خوشه را با یکدیگر محاسبه کرده و میانگین آن‌ها را به عنوان فاصله معیار در نظر می‌گیرد. نحوه کلی کار الگوریتم خوشه‌بندی سلسله مراتبی به صورت گام‌به‌گام به شرح زیر است:

گام ۱. شروع با مدوید n خوشه (مدوید نقاط توقف) و برچسب‌گذاری خوشه‌ها از ۱ تا n

گام ۲. محاسبه فاصله مدوید دو خوشه r و s با یکدیگر D(r, s) که در آن r, s = 1, 2, …, n هستند.

گام ۳. پیدا کردن خوشه‌هایی که مدوید آن‌ها کمتر از ۵۰۰ متر با یکدیگر فاصله دارند.

گام ٤. ادغام دو خوشه r و s که فاصله کمتر از ۵۰۰ متر با یکدیگر دارند و تولید خوشه جدید t.

گام ۵. تکرار گام۲ برای خوشه t و خوشه دلخواه k (k ≠ r, s)

گام ٦. تکرار گام ۳ و ٤

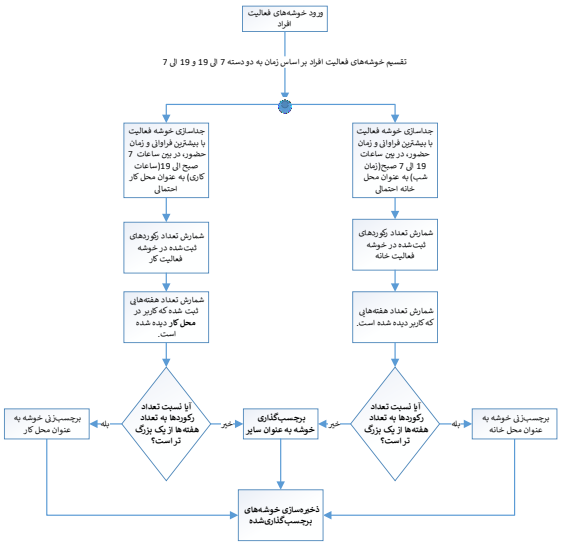
۲-۳ شناسایی هدف سفر

پس از شناسایی نقاط توقف و فعالیت افراد، نیاز به دانستن نوع فعالیت فرد در آن نقاط وجود دارد. به منظور بررسی نوع فعالیت افراد، طی سه مرحله موقعیت خانه، محل کار، و سایر فعالیتهای افراد شناسایی شد. با توجه به تفکیک‌پذیری مکانی پایین داده‌های تلفن همراه و در دست نداشتن نقشه‌های دقیق کاربری زمین در شهر تهران، شناسایی جزئی سایر اهداف سفر با خطای بالایی همراه است [Board et al., 2018] که از آن صرف نظر شد. در شناسایی موقعیت خانه افراد نیز به دلیل دقت مکانی پایین داده‌ها و کاربری متنوع شهری از روش پنجره زمانی استفاده شده است. بدین ترتیب خانه، موقعیتی در نظر گرفته شده است که فرد در روزهای کاری از ساعت ۱۹ الی ۷، و در روزهای غیرکاری در تمام ساعات روز بیش‌ترین تعداد مشاهده را در آن نقطه داشته است. با استفاده از الگوریتمی مشابه به همین

الگوریتم، محل کار به موقعیتی اتلاق شد که فرد در روزهای کاری از ساعت ۷ الی ۱۹، بیشترین حضور را در آنجا داشته است. در صورت شناسایی دو نقطه به عنوان محل کار، نقطه دورتر نسبت به خانه به عنوان محل کار در نظر گرفته شد، چرا که مطالعات نشان می‌دهند از بین نقاط پرمراجعه، نقطه‌ای که نسبت به خانه فاصله بیشتری دارد، معمولا محل کار فرد است [Levinson & Kumar, 1994; Schafer, 2000]. سایر نقاطی که نه به عنوان خانه و نه محل کار شناسایی شده‌اند، به عنوان محل انجام فعالیت‌های غیرکاری در نظر گرفته شدند. نکته دیگر آن است که نقطه‌ای به عنوان محل خانه یا کار فرد در نظر گرفته شد که فرد به طور میانگین حداقل یک بار در هفته در آن نقطه دیده شده باشد. در غیر این صورت، نقطه موردنظر به عنوان سایر نقاط فعالیتی در نظر گرفته شده است. روند کلی شناسایی اهداف سفر در شکل (۸) به صورت فلوچارت نمایش داده شده است.

۳-۳ برآورد ماتریس نمونه مبدا-مقصد و تولید-جذب

ماتریس مبدا-مقصد یکی از پرکاربردترین خروجی‌های مطالعات حمل‌ونقلی است که قادر به نمایش تبادل سفر بین مناطق مختلف می‌باشد. جهت استخراج ماتریس مبدا-مقصد، لازم است سفرهای هر فرد در طول روز شناسایی شوند. بازه زمانی در نظرگرفته‌شده به عنوان یک روز، از ساعت ۳ بامداد تا ۲:۵۹ روز بعد، لحاظ شد (روز موثر). علت آن نیز سفرهای شبانه است که ممکن است قبل از نیمه شب آغاز شده، و بعد از نیمه شب به پایان برسند. در این حالت، اگر نیمه‌شب به عنوان مبدا یک روز در نظر گرفته شود، سفرهایی بی‌پایان و بی‌آغاز در دو روز متوالی ثبت خواهند شد. لازم به ذکر است که در این مطالعه فرض شده است، تمام افراد صبح از خانه سفر خود را آغاز کرده و شب نیز به خانه باز می‌گردند [Alexander et al., 2015].



شکل ۸. فلوچارت شناسایی هدف سفر

پس از شناسایی سفرهای روزانه افراد، به شناسایی نوع و هدف سفر پرداخته شده است. با توجه به نوع فعالیت صورت گرفته در مبدا و مقصد سفر افراد (که در مرحله قبل تعیین شده است)، سفرهای هر فرد به سه دسته خانه‌مبنای کاری، خانه‌مبنای غیرکاری، و هیچ‌سرخانه تقسیم شدند. سپس ماتریس‌های مبدا-مقصد و تولید-جذب، به تفکیک اهداف سفر استخراج شدند. از سوی دیگر، اطلاعات تماس برای هر فرد در بازه زمانی یک‌ماهه موجود است. از این رو، جهت استخراج ماتریس مبدا-مقصد روزانه، میانگین تعداد سفرهای هر فرد در این بازه در نظر گرفته شده است. همچنین به منظور بررسی الگوی جابجایی در روزهای مختلف هفته، میانگین تعداد سفرهای هر فرد در هر روز در هفته در بازه زمانی یک‌ماهه مدنظر قرار گرفته است.

۳-٤ روش تعمیم ماتریس مبدا-مقصد نمونه به جامعه

همانطور که پیش‌تر اشاره شد، داده‌های مورد استفاده در این مطالعه، یک نمونه تقریباً ٤۰۰ هزار نفری از جامعه مورد بررسی، یعنی شهر تهران، هستند. برای اظهار نظر درباره کل جامعه، باید ماتریس مبدا-مقصد نمونه را به جامعه تعمیم داد. روش‌های گوناگونی برای تعمیم ماتریس نمونه به جامعه موجود هستند که بسته به دقت و جامعیت اطلاعات در دسترس، با یکدیگر تفاوت

دارند. با توجه به اینکه داده‌های موجود، فاقد اطلاعاتی از قبیل سن و جنسیت هستند، نمی‌توان به تفکیک سن، جنسیت، و یا سایر متغیرهای اقتصادی-اجتماعی، ضریب نفوذ تلفن همراه را محاسبه کرد [Imai et al., 2021]. در این مطالعه از روشی ساده اما مطمئن، برای تعمیم سفرها استفاده شده است. در این روش برای هر شخص در محدوده مورد مطالعه، یک ضریب تعمیم منحصربه‌فرد که با استفاده از رابطه (۵) محاسبه می‌گردد، در نظر گرفته شده است.

$$\beta_i = \frac{P_g}{\sum_{1}^{Q_g} U_{h_g}} \quad (۵)$$

که در آن:

β_i = ضریب تعمیم مربوط به کاربر i،

P_g = جمعیت ناحیه g بر اساس سرشماری،

U_{hg} = تعداد کل کاربرانی که آنتن موقعیت خانه آن‌ها در ناحیه g شناسایی شده است، و

Q_g = تعداد کل آنتن‌هایی که در ناحیه g قرار دارند، می‌باشند.

سپس با تقسیم این ضریب تعمیم بر تعداد روزهایی که هر کاربر در آن ناحیه شناسایی شده است، می‌توان این ضریب را به صورت روزانه برای افراد، محاسبه کرد (رابطه (٦)).

$$\theta_{ik} = \frac{\beta_i}{K_i} \quad (٦)$$

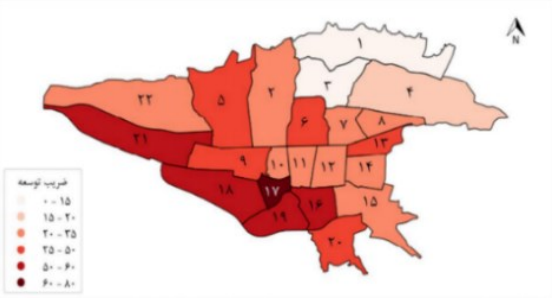
که در آن:

θ_{ik} = ضریب تعمیم روزانه کاربر i،

β_i = ضریب تعمیم کاربر i، و

K_i = تعداد روزهایی که حضور کاربر شناسایی شده است، می‌باشند.

در شکل (۹)، این ضرایب تعمیم بر روی نقشه شهر تهران، به تفکیک نواحی، نمایش داده شده است.



شکل ۹. توزیع ضرایب تعمیم سفر در شهر تهران به تفکیک نواحی به صورت روزانه

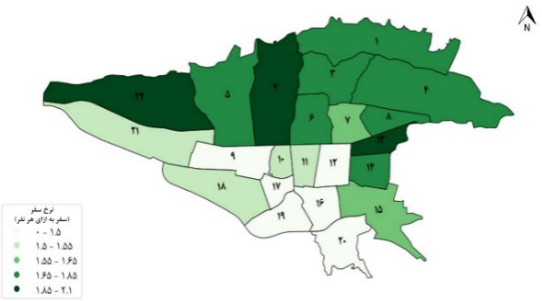
علت کمتر بودن این ضرایب برای نواحی شمال شهر نسبت به جنوب شهر را می‌توان به ضریب نفوذ بالاتر تلفن همراه در این نواحی، به دلایل اقتصادی، مرتبط دانست.

٤. ارائه و اعتبارسنجی نتایج

در این بخش نتایج مربوط به تولید و جذب سفر، نرخ سفر و ماتریس‌های مبدا-مقصد (تولید-جذب) در سطح نواحی ۲۲گانه ارائه شده است. با برآورد این مقادیر، امکان مقایسه با روش‌های بیشترپذیرفته‌شده مانند مطالعات جامع شهری فراهم آمد.

٤-۱ برآورد نرخ سفر به تفکیک مناطق ۲۲گانه

نرخ سفر یکی از خروجی‌های ارزشمند مطالعات حمل‌ونقلی است که از تقسیم تعداد سفرهای تولیدشده بر جمعیت ساکنان هر منطقه، به صورت روزانه محاسبه می‌گردد. با میانگین‌گیری از نرخ سفر تمام روزها در بازه مطالعه، میانگین نرخ روزانه سفر در شهر تهران بدست آمده است (شکل (۱۰)). به این ترتیب میانگین نرخ سفر در شهر تهران ۱,٦٤ محاسبه گردیده است، که در مقایسه با نرخ سفر ۱,۷، برآوردشده با استفاده از آماربرداری سفر خانوار (منبع: معاونت حمل و نقل و ترافیک شهرداری تهران)، اختلاف اندکی دارد. این اختلاف را می‌توان ناشی از عدم تشخیص برخی سفرهای کوتاه، بخصوص سفرهای با طول کمتر از ۵۰۰ متر، توسط داده‌های تلفن همراه دانست.

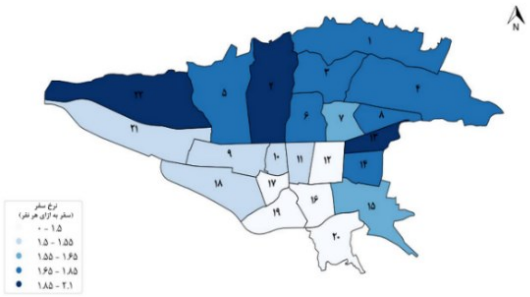


شکل ۱۰. میانگین نرخ روزانه سفر در شهر تهران به تفکیک نواحی

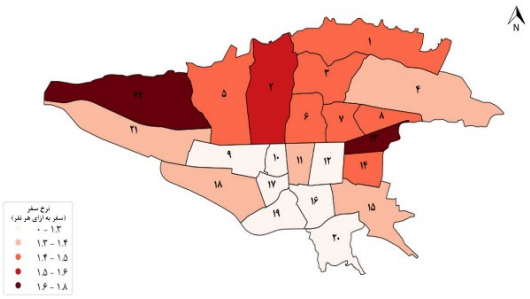
همانطور که ملاحظه می‌گردد، نرخ سفر در نواحی شمال تهران، به طور معناداری از نواحی جنوب شهر بیشتر است. این تفاوت را می‌توان ناشی از احتمال بیشتر نرخ بالای اشتغال در شمال شهر و اشتغال بیش از یک نفر در هر خانوار، کوچکتر بودن بعد خانوار با توجه به تفاوت وضعیت معیشت و سبک زندگی، و احتمال نزدیکی محل سکونت و کار در جنوب شهر دانست.

به منظور مطالعه دقیق‌تر، نرخ سفرها به تفکیک روزهای کاری و غیرکاری نیز بررسی شدند که نتایج آن به ترتیب در شکل‌های (۱۱) و (۱۲) قابل‌مشاهده است. میانگین نرخ سفر در روزهای کاری ۱,٦٤، و در روزهای غیرکاری ۱,٤۰، محاسبه شده است.

با دقت در نرخ سفر، نکات جالب‌توجهی قابل برداشت است. به عنوان مثال، علاوه بر نرخ کمتر کلی سفر در مناطق جنوب شهر، نرخ سفرهای غیرکاری در جنوب شهر، بسیار کمتر از شمال شهر است. این موضوع را می‌توان به سطح پایین درآمد افراد در نواحی جنوب شهر و در نتیجه نرخ کمتر سفرهای غیرضروری و تفریحی مرتبط دانست و در مطالعات مرتبط با عدالت اجتماعی، مدنظر قرار داد. لازم به ذکر است با توجه به این که آماربرداری سفر در مطالعات طرح جامع تنها در یک روز نمونه کاری صورت می‌گیرد، نرخ سفر به تفکیک روزهای کاری و غیرکاری از این مطالعات قابل‌استخراج و قابل مقایسه با نرخ‌های سفر به‌دست‌آمده از داده‌های تلفن همراه نیست.



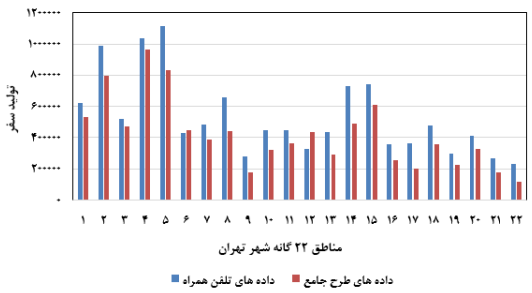
شکل ۱۱. میانگین نرخ روزانه سفر در شهر تهران به تفکیک نواحی برای روزهای کاری



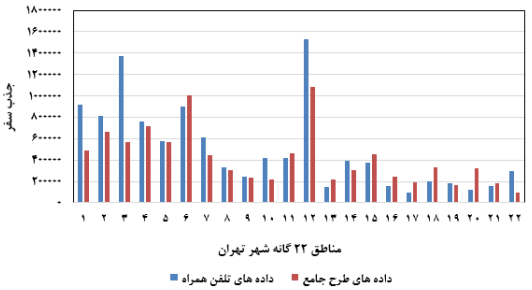
شکل ۱۲. میانگین نرخ روزانه سفر در شهر تهران به تفکیک نواحی برای روزهای غیرکاری

٤-۲ برآورد تولید-جذب سفر به تفکیک مناطق ۲۲گانه

یکی دیگر از خروجی‌هایی که می‌تواند به منظور اعتبارسنجی داده‌های تلفن همراه مورد استفاده قرار گیرد، تولید و جذب سفر در سطح مناطق شهر تهران است. پس از برآورد سفرها و تعمیم آن به جامعه، با استفاده از تجمیع سفرهای تولیدشده و جذب‌شده در سطح مناطق، تولید-جذب سفرها به تفکیک مناطق ۲۲گانه شهرداری تهران استخراج و نتایج آن با مطالعات جامع حمل‌ونقل تهران مقایسه شد (شکل (۱۳) و (۱٤)).



شکل ۱۳. مقایسه تولید سفر با استفاده از داده‌های تلفن همراه و مطالعات جامع حمل و نقل شهر تهران



شکل ۱٤. مقایسه جذب سفر با استفاده از داده‌های تلفن همراه و مطالعات جامع حمل و نقل شهر تهران

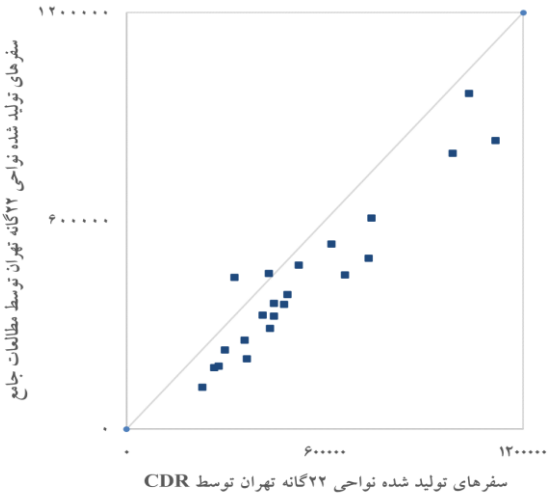
همانطور که مشاهده می‌شود، تولید–جذب به‌دست‌آمده از داده‌های تلفن همراه و مطالعات جامع حمل و نقل شهر تهران در هر دو روش به یکدیگر نزدیک هستند. البته در مواردی مانند منطقه ۳ نتایج حاصل از داده‌های تلفن همراه میزان جذب بیشتری را نسبت به مطالعات جامع، نشان می‌دهند که در ادامه به بررسی این تفاوت‌ها پرداخته می‌شود.

در منطقه ۳ در شهر تهران نسبت به زمان برداشت آمارهای مربوط به مطالعات طرح جامع، کاربری‌های جاذب سفر بسیاری در این ناحیه از جمله مجتمع‌های جدید و متعدد تجاری اداری، پارک آب و آتش، مجموعه باغ کتاب و کتابخانه ملی ایران ایجاد شده است. میزان جذب بالای این منطقه را می‌توان با نرخ بالای توسعه کاربری‌های اداری–تجاری و تفریحی در این منطقه نسبت به سایر مناطق توجیه کرد؛ چرا که مطالعات جامع در بازه‌های نسبتاً طولانی به‌روزرسانی می‌شوند و این تغییرات در آن‌ها دیده نمی‌شوند. با این حال، بررسی دقت این مقدار جذب حاصل از داده‌های تلفن همراه، نیازمند مطالعات بیشتر است. این مسئله، خود بیانگر یکی از مزایای استفاده از داده‌های تلفن همراه، به عنوان مکمل مطالعات جامع است. همچنین با دقت در شکل (۱٤)، رشد جذب سفر در تعدادی دیگر از مناطق، از جمله ۱، ۱۰، ۱۲، و ۲۲ نیز مشهود است. مشابه منطقه ۳، علت این تفاوت‌ها را تا حدودی می‌توان به رشد کاربری‌های اداری–تجاری، تفریحی، و مسکونی در این مناطق، ربط داد. به عنوان مثال منطقه ۲۲ که با بهره‌برداری دریاچه مصنوعی چیتگر و بازار بزرگ ایران همراه بوده است، خود دلیلی بر افزایش جذب سفر به این منطقه است.

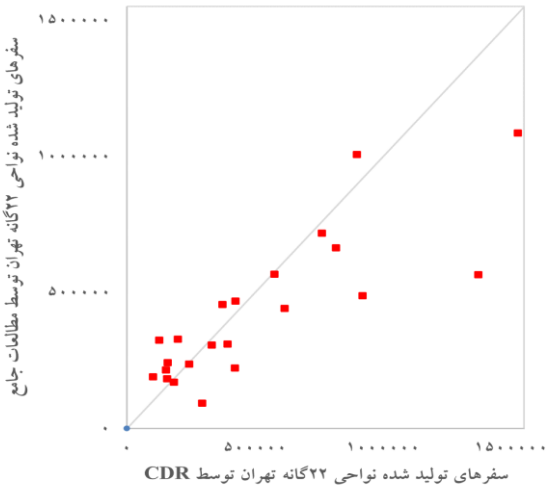
علاوه بر اعتبارسنجی منطقی، همبستگی میان تولید–جذب حاصل از داده‌های تلفن همراه و مطالعات جامع مورد بررسی قرار گرفت. به منظور بررسی این همبستگی از ضریب همبستگی پیرسون یا همان ρ استفاده شده است. برای نمایش بهتر، نمودار پراکنشی دو مقدار متناظر برای تولید و جذب ترسیم شده است (شکل (۱۵) و (۱٦)).

برای تولید سفر ρ = ۰٫۹۵ و برای جذب سفر ρ = ۰٫۸۳ محاسبه می‌گردد. همانطور که مشاهده می‌شود همبستگی تولید سفر نسبت به جذب سفر، بالاتر بوده که می‌تواند نشان‌دهنده تطابق بیشتر داده‌های تلفن همراه با مطالعات جامع در تولید سفر، نسبت به جذب سفر باشد. در مطالعات پیشین بررسی‌شده، مقادیر تولید–جذب به طور مجزا مقایسه نشده‌اند [Board et al., 2018]. البته در برخی مطالعات، مقایسه بین کل سفرهای تولیدشده و جذب‌شده انجام شده است. جولاک و همکاران در سال ۲۰۱۵ ضریب همبستگی سفرهای ایجادشده برای شهر ریو را ۰٫۸۱ [Çolak et al., 2015] اعلام می‌کنند. لندمارک و همکاران نیز مقدار ρ = ۰٫۹۱ را برای کل سفرها اعلام کرده‌اند [Dypvik Landmark et al., 2021] این مقدار برای مطالعه حاضر ۰٫۹۳ می‌باشد که منطقی به نظر می‌رسد. مقدار پایین‌تر ضریب همبستگی جذب سفر در مقایسه با تولید را می‌توان با تغییرات زیاد کاربری‌های اداری‌تجاری و تفریحی در

مناطق مختلف تهران، همانطور که پیش‌تر نیز اشاره شد، مرتبط دانست.



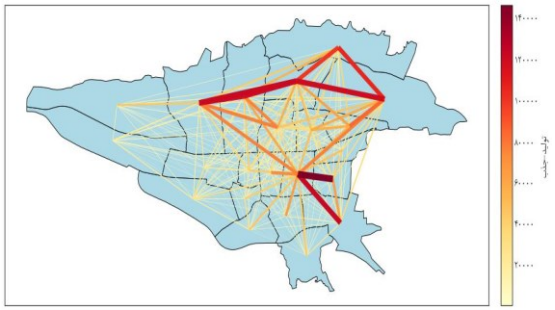
شکل ۱۵. سنجش همبستگی تولید حاصل از داده‌های CDR و مطالعات جامع



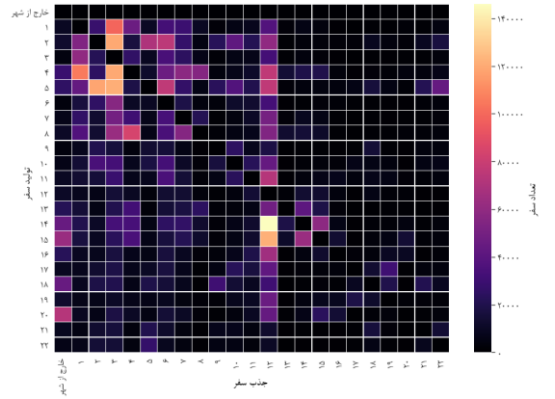
شکل ۱۶. سنجش همبستگی جذب حاصل از داده‌های CDR و مطالعات جامع

۳-٤ ماتریس برآوردشده تولید-جذب و شناسایی الگوهای جابجایی بین مناطق ۲۲گانه شهر تهران

ماتریس تولید-جذب [مبدا-مقصد] به خوبی امکان بررسی توزیع سفر بین مناطق را فراهم می‌کند. این ماتریس‌ها به تفکیک روزهای کاری و غیرکاری و بسته به اهداف سفر استخراج شده‌اند. نمونه‌ای از ماتریس تولید-جذب سفرهای خانه‌مبنای کاری در روزهای کاری در قالب خطوط تمایل و ماتریس حرارتی، قابل مشاهده است (شکل‌های (۱۷) و (۱۸)) لازم به ذکر است ماتریس‌ها به تفکیک سایر اهداف سفر و یا سایر روزهای هفته نیز قابل ارائه است و در این بخش تنها نمونه‌ای از ماتریس‌های استخراج‌شده ارائه شده‌اند.



شکل ۱۷. خطوط مطلوبیت تولید-جذب سفرهای خانه‌مبنای کاری در روزهای کاری

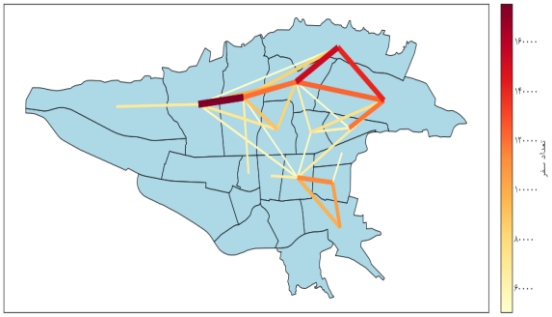


شکل ۱۸. ماتریس حرارتی تولید-جذب سفرهای خانه‌مبنای کاری در روزهای کاری

در ماتریس حرارتی (شکل (۱۸)) رنگ‌های روشن‌تر نشانه بیشتر بودن تعداد سفرهای مبادله‌شده بین مناطق هستند. اما در خطوط تمایل (شکل (۱۷)) رنگ‌های تیره نمایانگر تعداد سفر بیشتر هستند. همانطور که ملاحظه می‌گردد، بیشینه سفرها به ترتیب از منطقه ١٤ و ١٥ به منطقه ١٢ صورت گرفته است. همانطور که پیش‌تر هم بیان شد، منطقه ١٢ به دلیل کاربری‌های متراکم تجاری موجود در آن، از جذابیت بالایی برخوردار است. مناطق ١٤ و ١٥ نیز شامل محله‌های افسریه، مسعودیه، شوش، و خاوران است که از بافت مسکونی بسیار متراکمی برخوردارند و تولید سفر

زیاد از این نواحی منطقی به نظر می‌رسد. منطقه ۳ نیز جاذب سفرهای زیادی است که پیش‌تر به دلایل آن پرداخته شد.

به منظور بررسی دقیق‌تر تبادل سفرها در این مطالعه، پنجاه خط مطلوبیت با بیشینه تقاضا استخراج و ترسیم شده‌اند. شکل (۱۹) خطوط مطلوبیت برای تمامی اهداف سفر در روزهای کاری را نشان می‌دهد.



شکل ۱۹. خطوط مطلوبیت با بیشینه تقاضا برای تمام سفرها در روزهای کاری

با توجه به شکل (۱۹)، می‌توان گفت بخش قابل توجهی از سفرهای شهر تهران در سه خوشه کلی صورت می‌گیرند؛ سفرهای بین شمال‌غرب و مرکز تهران، سفرهای بین شمال‌شرق و مرکز تهران، و سفرهای بین جنوب شرق و منطقه ۱۲ تهران. همان‌طور که پیش‌تر نیز بیان شد، این نتایج می‌تواند بیانگر تمرکز بالای کاربری‌های اداری-تجاری در مناطق ۳، ٦ (مرکز تهران)، و ۱۲ باشد. همچنین تمرکز بالای مناطق مسکونی در شمال و شرق و همچنین برخی نواحی جنوب شرق تهران را می‌توان علت تبادل سفر بالای این مناطق دانست. به نظر می‌رسد نتایج حاصل، با رفتار سفر افراد همخوانی مناسبی داشته‌اند.

تبادلات بالای سفر بین نواحی ۲ و ۵، نواحی ۱، ۳ و ٤ و نواحی ٤ و ۸ نیز درخور توجه است. این موضوع می‌تواند در توسعه خطوط حمل‌ونقل همگانی مورد توجه قرار گیرد. همچنین با توجه به مجاورت نواحی مذکور، ساخت و توسعه خطوط ویژه دوچرخه‌سواری نیز می‌تواند یک گزینه دیگر باشد. مشابه این الگو، بین نواحی ۱۲، ۱٤، ۱۵، و ۱٦ نیز مشهود است. هر کدام از این گزینه‌های پیشنهادی می‌تواند با امکان‌سنجی اقتصادی مناسب، برای این مناطق نیز کارآمد باشد.

۵. جمع‌بندی و پیشنهادات

در این مطالعه با استفاده از داده‌های تلفن همراه، رفتار سفر افراد در شهر تهران مورد بررسی قرار گرفت. بدین منظور، با توجه به ماهیت کلان پایگاه داده موجود، ابتدا پیش‌پردازش لازم بر روی داده‌ها صورت گرفت. در این راستا، پس از پالایش اولیه داده‌ها که شامل حذف داده‌های پرت، اصلاح ساختاری داده‌ها و اصلاح نمونه، به بررسی تفکیک‌پذیری زمانی و مکانی آن‌ها پرداخته شد. سپس با استفاده از الگوریتم‌های یادگیری ماشین، نقاط توقف و فعالیت افراد شناسایی گردیده و اهداف سفر آن‌ها دسته‌بندی شد. با شناسایی سفرهای هر فرد در طول روز، امکان استخراج ماتریس مبدا-مقصد برای نمونه داده‌های تلفن همراه فراهم گردید. سپس ماتریس نمونه با محاسبه ضرایب تعمیم بر اساس جمعیت برآوردشده توسط داده‌های تلفن همراه و جمعیت براساس سرشماری، به ماتریس جامعه تبدیل شد. در نهایت، با در دست داشتن این ماتریس، امکان مشاهده نحوه تبادل سفر بین مناطق مختلف، میسر شد. همچنین نرخ سفر و تولید-جذب سفر به تفکیک مناطق ۲۲گانه شهر تهران ارائه و نتایج حاصل با مطالعات جامع حمل‌ونقلی شهر تهران مقایسه گردید. در این مطالعه، با استفاده از خروجی حاصل از تحلیل داده‌های تلفن همراه، نشان داده شد که این نوع داده‌ها از توانایی بالایی در برآورد بسیاری از موارد مدنظر در مطالعات حمل‌ونقل برخوردار هستند؛ چرا که نرخ سفر در سطح شهر تهران، و مقادیر تولید-جذب اختلاف اندکی با خروجی‌های متناظر مطالعات جامع داشتند. بنابراین پیشنهاد می‌شود در مسائل مختلف برنامه‌ریزی حمل‌ونقل استفاده بیشتری از داده‌های تلفن همراه گردد.

داده‌های طرح جامع در یک روز خاص از سال بدست می‌آیند. مطالعه نشان داد که الگوی توزیع سفرها برای یک روز خاص تطبیق مناسبی ندارند. اما چون این نوع داده‌ها برای روزها و ماه‌های مختلف سال با جزئیات به مراتب بیشتری است، بنابراین

می‌توان این الگوها را برای بازه‌های زمانی مختلف و با جزئیات بیشتری در جهت برنامه‌ریزی در اختیار داشت.در نهایت به نظر می‌رسد داده‌های تلفن همراه، علاوه بر محاسبه شاخص‌های مختلف حمل‌ونقلی، می‌توانند در مسائل برنامه‌ریزی حمل‌ونقل، مانند بحث‌های عدالت اجتماعی، تخصیص حمل و نقل همگانی، مطالعات زیست‌محیطی، و غیره، بکار گرفته شوند. پیشنهاد می‌شود در مطالعات آتی بر روی این داده‌ها، ضمن افزایش دقت زمانی و مکانی آن‌ها، در زمینه‌های مختلف برنامه‌ریزی بکار گرفته شده و اعتبارسنجی شوند، تا بتوان مانند روش‌های تکامل‌یافته کنونی، از آن‌ها به عنوان مکمل برای آماربرداری‌های سنتی استفاده کرد.

٦. پی‌نوشت‌ها

1. Call Detail Records
2. Tower-based CDR
3. Triangular CDR
4. Cellular Network Signaling Data
5. Location Area Update
6. Periodic Location Update
7. transient origin destination (t-OD) matrix
8. tower-to-tower t-OD
9. node-to-node t-OD
10. medoid
11. Ping-Pong Handover
12. Voronoi
13. Stop points
14. Stay points
15. Haversine
16. Longitude
17. Latitude
18. Handover
19. Hierarchical Agglomerative Clustering (HAC)
20. Metrics
21. Linkage Criteria

۷. منابع

– اربابی بیگدلی، محمد(۱۳۹۷) " برآورد ماتریس تقاضای مبدا–مقصد با داده‌های شمارش حجم و شبکه تلفن همراه (مطالعه موردی: شهر شیراز)" پایان‌نامه کارشناسی ارشد، استاد راهنما: هدایت ذکایی آشتیانی، تهران: دانشکده مهندسی عمران، گروه حمل‌ونقل، دانشگاه صنعتی شریف.

– افتخار ، زهرا (۱۳۹۷) " برآورد جدول تقاضای سفر مبدا–مقصد با استفاده از داده‌های مکانی–زمانی شبکه تلفن همراه گردآوری‌شده به روش PLU" پایان‌نامه کارشناسی ارشد، استاد راهنما: یوسف شفاهی، تهران: دانشکده مهندسی عمران، گروه حمل‌ونقل، دانشگاه صنعتی شریف.

– نوری ، میثم (۱۳۹۸) " برآورد نقشینه‌های شهروندان با استفاده از اطلاعات مبدا–مقصد و داده‌های شبکه تلفن همراه: مطالعه موردی شهر شیراز" پایان‌نامه کارشناسی ارشد، استاد راهنما: یوسف شفاهی، تهران: دانشکده مهندسی عمران، گروه حمل‌ونقل، دانشگاه صنعتی شریف.

– زاهدی ، مصطفی (۱۳۹۵) " برآورد الگوهای فعالیتی مسافران در شهر با استفاده از داده‌های مکانی–زمانی حاصل از شبکه تلفن همراه" پایان‌نامه کارشناسی ارشد، استاد راهنما: یوسف شفاهی، تهران: دانشکده مهندسی عمران، گروه حمل‌ونقل، دانشگاه صنعتی شریف.

- Alexander, L., Jiang, S., Murga, M., & González, M. C. (2015) “Origin–destination trips by purpose and time of day inferred from mobile phone data”, Transportation Research Part C: Emerging Technologies, Vol. 58, PP. 240-250.

- Barboza, M. H. C., Alencar, R. d. S., Chaves, J. C., Silva, M. A. H. B., Orrico, R. D., & Evsukoff, A. G. (2021) “Identifying Human Mobility Patterns in the Rio de Janeiro Metropolitan Area using Call Detail Records”, Transportation Research Record, Vol. 2675, No. 4, pp. 213-221.

- Board, T. R., National Academies of Sciences, E., & Medicine. (2018) “Cell Phone Location Data for Travel Behavior Analysis”, Washington, DC: The National Academies Press.

- Bonnel, P., Fekih, M., & Smoreda, Z. (2018) “Origin-Destination estimation using mobile network probe data”, Transportation Research Procedia, Vol. 32, pp. 69-81.

- Caceres, N., Wideberg, J. P., & Benitez, F. G. (2008) “Review of traffic data estimations extracted from cellular networks”, IET Intelligent Transport Systems, Vol. 2, No. 3, pp. 179-192.

- Calabrese, F., Colonna, M., Lovisolo, P., Parata, D., & Ratti, C. (2011) “Real-Time Urban Monitoring Using Cell Phones: A Case Study in Rome. Intelligent Transportation Systems”, IEEE Transactions on, Vol. 12, pp. 141-151.

- Calabrese, F., Di Lorenzo, G., Liu, L., & Ratti, C. (2011) “Estimating Origin-Destination flows using opportunistically collected mobile phone location data from one million users in Boston Metropolitan Area”.

- Chen, C., Ma, J., Susilo, Y., Liu, Y., & Wang, M. (2016) “The promises of big data and small data for travel behavior (aka human mobility) analysis”, Transportation Research Part C: Emerging Technologies, Vol. 68, pp. 285-299.

- Chen, G., Viana, A. C., Fiore, M., & Sarraute, C. (2019) “Complete trajectory reconstruction from sparse mobile phone data”, EPJ Data Science, Vol. 8, No. 1, p. 30.

- Chin, K., Huang, H., Horn, C., Kasanicky, I., & Weibel, R. (2019) “Inferring fine-grained transport modes from mobile phone cellular signaling data”, Computers, Environment and Urban Systems, Vol. 77.

- Cintia, P., Pappalardo, L., Rinzivillo, S., Fadda, D., Boschi, T., Giannotti, F., et al. (2020) “The relationship between human mobility and viral transmissibility during the COVID-19 epidemics in Italy”, arXiv preprint arXiv:2006.03141.

- Çolak, S., Alexander, L. P., Alvim, B. G., Mehndiratta, S. R., & González, M. C. (2015) “Analyzing Cell Phone Location Data for Urban Travel: Current Methods, Limitations, and Opportunities”, Transportation Research Record, Vol. 2526, No. 1, pp. 126-135.

- Dypvik Landmark, A., Arnesen, P., Södersten, C.-J., & Hjelkrem, O. A. (2021) “Mobile phone data in transportation research: methods for benchmarking against other data sources”, Transportation, Vol. 48, No. 5, pp. 2883-2905.

- Fekih, M., Bellemans, T., Smoreda, Z., Bonnel, P., Furno, A., & Galland, S. (2020) “A data-driven approach for origiŽ destination matrix construction from cellular network signalling data: a case study of Lyon region (France)”, Transportation, pp. 1-32.

- Ghurye, J., Krings, G., & Frias-Martinez, V. (2016, 13-16 June 2016) “A Framework to Model Human Behavior at Large Scale during Natural Disasters”, The 2016 17th IEEE International Conference on Mobile Data Management (MDM).

- González, M. C., Hidalgo, C. A., & Barabási, A.-L. (2008) “Understanding individual human mobility patterns”, Nature, Vol. 453, No. 7196, pp. 779-782.

- Grantz, K. H., Meredith, H. R., Cummings, D. A. T., Metcalf, C. J. E., Grenfell, B. T., Giles, J. R., et al. (2020) “The use of mobile phone data to inform analysis of COVID-19 pandemic epidemiology”, Nature Communications, Vol. 11, No. 1, pp. 4961.

- Hankaew, S., Phithakkitnukoon, S., Demissie, M. G., Kattan, L., Smoreda, Z., & Ratti, C. (2019) "Inferring and modeling migration flows using mobile phone network data", IEEE Access, Vol. 7, pp. 164746-164758.

- Hernandez, M., Hong, L., Frias-Martinez, V., Whitby, A., & Frias-Martinez, E. (2017) "Estimating poverty using cell phone data: evidence from Guatemala", World Bank Policy Research Working Paper, No. 7969.

- Hoteit, S., Secci, S., Sobolevsky, S., Ratti, C., & Pujolle, G. (2014) "Estimating human trajectories and hotspots through mobile phone data", Comput. Netw., Vol. 64, pp. 296–307.

- Imai, R., Ikeda, D., Shingai, H., Nagata, T., & Shigetaka, K. (2021) "Origin-Destination Trips Generated from Operational Data of a Mobile Network for Urban Transportation Planning", Journal of Urban Planning and Development, Vol. 147, No. 1.

- Iqbal, M. S., Choudhury, C. F., Wang, P., & González, M. C. (2014) "Development of origin–destination matrices using mobile phone call data", Transportation Research Part C: Emerging Technologies, Vol. 40, pp. 63-74.

- Janecek, A., Hummel, K. A., Valerio, D., Ricciato, F., & Hlavacs, H. (2012) "Cellular data meet vehicular traffic theory: location area updates and cell transitions for travel time estimation", The Proceedings of the 2012 ACM Conference on Ubiquitous Computing, Pittsburgh, Pennsylvania.

- Jiang, S., Ferreira, J., & Gonzalez, M. C. (2017) "Activity-Based Human Mobility Patterns Inferred from Mobile Phone Data: A Case Study of Singapore", IEEE Transactions on Big Data, 3(2), 208-219.

- Jiang, W., & Luo, J. (2022) "Big Data for Traffic Estimation and Prediction: A Survey of Data and Tools", Applied System Innovation, 5(1), 23.

- Jing, W., Dianhai, W., Xianmin, S., & Di, S. (2011, 28-29 March 2011) "Dynamic OD Expansion Method Based on Mobile Phone Location", Paper presented at the 2011 Fourth International Conference on Intelligent Computation Technology and Automation.

- Levinson, D. M., & Kumar, A. (1994) "The Rational Locator: Why Travel Times Have Remained Stable", Journal of the American Planning Association, 60(3), 319-332.

- Lwin, K. K., Sekimoto, Y., & Takeuchi, W. (2018) "Estimation of Hourly Link Population and Flow Directions from Mobile CDR", ISPRS International Journal of Geo-Information, 7(11), 449.

- Mamei, M., Bicocchi, N., Lippi, M., Mariani, S., & Zambonelli, F. (2019) "Evaluating Origin–Destination Matrices Obtained from CDR Data", Sensors, 19(20), 4470.

- Pourmoradnasseri, M., Khoshkhah, K., Lind, A., & Hadachi, A. (2019, 21-23 Oct. 2019) "OD-Matrix Extraction based on Trajectory Reconstruction from Mobile Data", Paper presented at the 2019 International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob).

- Qin, S., Man, J., Wang, X., Li, C., Dong, H., & Ge, X. (2019) "Applying Big Data Analytics to Monitor Tourist Flow for the Scenic Area Operation Management", Discrete Dynamics in Nature and Society, 2019, 1-11.

- Sakamanee, P., Phithakkitnukoon, S., Smoreda, Z., & Ratti, C. (2020) "Methods for Inferring Route Choice of Commuting Trip From Mobile Phone Network Data", ISPRS International Journal of Geo-Information, Vol. 9, No. 5, P. 306.

- Schafer, A. (2000) “Regularities in travel demand: an international perspective”.

- Schneider, C. M., Belik, V., Couronné, T., Smoreda, Z., & González, M. C. (2013) “Unravelling daily human mobility motifs”, Journal of The Royal Society Interface, Vol. 10, No. 84.

- Shokri, V., & Abbaspour, R. A. (2017) “Using Map Matching Algorithms to Extract Traffic Information from Low Sampling Rate GPS Trajectories”, Quarterly Journal of Transportation Engineering, Vol. 8, No. 4, pp. 529-544.

- Sikder, R., Uddin, M. J., & Halder, S. (2016, 12-13 Dec. 2016) “An efficient approach of identifying tourist by call detail record analysis”, The 2016 International Workshop on Computational Intelligence (IWCI).

- Toole, J. L., Colak, S., Sturt, B., Alexander, L. P., Evsukoff, A., & González, M. C. (2015) “The path most traveled: Travel demand estimation using big data resources”, Transportation Research Part C: Emerging Technologies, Vol. 58, pp. 162-177.

- Vanhoof, M., Reis, F., Smoreda, Z., & Ploetz, T. (2018) “Detecting home locations from CDR data: introducing spatial uncertainty to the state-of-the-art”, arXiv preprint arXiv:1808.06398.

- Wang, C., & Hess, D. B. (2021) “Role of Urban Big Data in Travel Behavior Research”, Transportation Research Record, Vol. 2675, No. 4, pp. 222-233.

- Wang, P., Hunter, T., Bayen, A. M., Schechtner, K., & González, M. C. (2012) “Understanding Road Usage Patterns in Urban Areas”, Scientific Reports, Vol. 2, No. 1, p. 1001.

مهدی‌یار درویش‌زاده درجه کارشناسی در رشته مهندسی عمران را در سال ۱۳۹۷ از دانشگاه آزاد مشهد اخذ نمود و در حال حاضر دانشجوی کارشناسی ارشد مهندسی و برنامه‌ریزی حمل‌ونقل در دانشگاه علم و صنعت ایران است. زمینه های پژوهشی مورد علاقه ایشان برنامه‌ریزی حمل‌ونقل، بررسی رفتار سفر، و کار با کلان‌داده‌ها است.

سعید رحمانی درجه کارشناسی مهندسی عمران را در سال ۱۳۹۳ از دانشگاه علم و صنعت ایران و درجه کارشناسی ارشد در رشته مهندسی و برنامه ریزی حمل‌ونقل را در سال ۱۳۹٦ از همان دانشگاه اخذ نمود. زمینه‌های پژوهشی مورد علاقه ایشان مهندسی ترافیک، سیستمهای حمل‌ونقل هوشمند، کلان‌داده، داده‌کاوی، مدل‌های رفتاری بوده و در حال حاضر کارشناس ارشد مرکز تحلیل داده‌های حمل‌ونقل دانشگاه علم و صنعت ایران است.

افشین شریعت مهیمنی درجه کارشناسی در مهندسی عمران را در سال ۱۳٦۹ از دانشگاه علم و صنعت ایران و درجه کارشناسی ارشد در رشته مهندسی و برنامه‌ریزی حمل‌ونقل را در سال ۱۳۷٤ از همان دانشگاه اخذ نمود. در سال ۱۳۸۰ موفق به کسب درجه دکتری در رشته مهندسی عمران از دانشگاه علم و صنعت ایران گردید. زمینه‌های پژوهشی مورد علاقه ایشان مهندسی ترافیک، حمل‌ونقل همگانی، قابلیت اطمینان، شبیه سازی ترافیکی وتحلیل داده و برنامه‌ریزی حمل‌ونقل بوده و در حال حاضر عضو هیات علمی با مرتبه استاد در دانشگاه علم و صنعت ایران است.

سینا عبداللهی درجه کارشناسی در رشته مهندسی عمران را در سال ۱۳۹۹ از دانشگاه صنعتی امیرکبیر اخذ نمود و در حال حاضر دانشجوی کارشناسی ارشد مهندسی و برنامه‌ریزی حمل‌ونقل در دانشگاه علم و صنعت ایران است. زمینه های پژوهشی مورد علاقه ایشان برنامه‌ریزی حمل‌ونقل، بررسی رفتار سفر، و کار با کلان‌داده‌ها است.