



ارائه شده توسط:

سایت ترجمه فا

مرجع جدیدترین مقالات ترجمه شده

از نشریات معتبر

الگوهای تحرک انسانی مبتنی بر فعالیت استنباط شده از داده های تلفن همراه:

یک مطالعه موردی از سنگاپور

چکیده: در این مطالعه، کشور سنگاپور را بعنوان یک مثال در نظر می‌گیریم، نشان می‌دهیم که چگونه می‌توانیم داده‌های حاوی سوابق جزئیات تماس تلفن همراه (CDR) را که حاوی میلیون‌ها کاربر ناشناس است استفاده کنیم، تا شبکه‌های منحصربفرد قابل قیاس با رویکرد مبتنی بر فعالیت را استخراج کنیم. چنین رویکردی بطور وسیع در عمل برنامه‌ریزی حمل‌ونقل به منظور توسعه میکرو شبیه‌سازی‌های شهری از فعالیت‌های روزانه فردی و مسافرت استفاده می‌شود؛ با اینحال شدیداً به داده‌های دقیق بررسی سفر بستگی دارد تا رفتار مبتنی بر فعالیت منحصربفرد را به دست آورد. یک چارچوب داده‌کاوی نوآورانه را ارائه می‌کنیم که تکنیک‌های پیشرفته تر در استخراج الگوهای تحرک از داده‌های خام CDR تلفن همراه را ترکیب می‌کند، و خط لوله‌ای را طراحی می‌کند که سوابق انبوه و منفعل تلفن همراه را به الگوهای معنادار تحرک انسانی فضایی برگردان می‌کند که به راحتی قابل تفسیر برای اهداف برنامه‌ریزی حمل‌ونقل و شهری است. با توجه به رشد روز افزون تلفن‌های همراه، و کاهش و کوچک شدن کار و منابع مالی در بخش‌های عمومی در سراسر جهان، روش ارائه شده در این تحقیق بعنوان یک جایگزین کم هزینه برای سازمان‌های حمل‌ونقل و برنامه‌ریزی مورد استفاده قرار می‌گیرد تا الگوهای فعالیت انسانی در شهرها را درک کند، و طرح‌های هدفمندی را برای توسعه پایدار آتی ارائه کند.

کلیدواژه: داده های تلفن همراه، داده های مسیریابی، شبکه های تحرک انسانی، تشخیص نقاط تحرک، محاسبات

شهری

1. مقدمه

برای بهبود تحرک، قابلیت دسترسی، و کیفیت زندگی، درک این موضوع که چگونه افراد مسافرت می‌کنند و فعالیت‌های خودشان را انجام می‌دهند، تمرکز اصلی برنامه‌ریزان شهری و حمل‌ونقل و جغرافیادانان می‌باشد [1]، [2]، [3]، [4].

در گذشته، این کار با جمع‌آوری داده‌های بررسی در نمونه‌های ساده کوچک و فرکانس‌های پایین (بعنوان مثال، سازمان‌های برنامه‌ریزی مناطق کلان‌شهر در کشورهای توسعه یافته که یک درصد از سفرهای خانوادگی را یک یا دو بار در یک دهه انجام می‌دهند) انجام می‌شد. با تکامل جامعه و نوآوری در فن‌آوری، شهرها در مقایسه با هر زمان دیگری در جهان متصل بهم متنوع‌تر و پیچیده‌تر شده‌اند. امروزه بیش از نیمی از جمعیت جهان (54 درصد در سال 2014) در مناطق شهری زندگی می‌کنند، و پیش‌بینی می‌شود که تا سال 2050 جمعیت ساکن در شهرها به 2.5 میلیارد نفر برسد [5]. روش‌های مرسوم که عمدتاً در حوزه برنامه‌ریزی-حمل‌ونقل کاربرد دارند به منظور وفق دادن و تناسب-سازی داده‌های کوچک و البته پرهزینه‌ی جمع‌آوری شده توسعه می‌یابند، و چالش‌های فعلی را برآورده نمی‌سازند. این موضوع برای محققان شهری ضروری است تا به منظور رفع چالش‌های شهری مانند ازدحام ترافیک، آلودگی محیط زیست و تخریب، و افزایش مصرف انرژی و انتشار گازهای گلخانه‌ای رویکردهای جدیدی را دنبال کنند. با ظهور فن‌آوری‌های سنجش همه جا حاضر، و همچنین ردپای انسان دیجیتال، که اثرات و علائم دیجیتالی است که مردم در زمان تعامل با فضاهای فیزیکی سایبری به جا می‌گذارند [6]، [7]، می‌توان در مقیاس‌های بدون محدودیت و وسیع همراه با فرکانس بالا و هزینه‌های پائین هر فعل و انفعالی را ثبت نمود. این فرصت بزرگی را برای تغییر چشم‌انداز در تئوری تحقیقات شهری برای یک افق جدید به ارمغان می‌آورد (بعنوان مثال، [8]، [9]، [10]، [11])، و مستلزم فن‌آوری برای پیوند به داده‌های عظیم با تئوری شهری است و همچنین نیازمند تفکری است که دانش جدید شهری، به نام محاسبات شهری یا علم نوین شهری را به دست آورد [7]، [12]، [13].

این مقاله کاربرد تحلیل‌های داده‌های بزرگ را نشان می‌دهد که داده‌های تلفن همراه همه جا حاضر به الگوهای تحرک انسانی قابل تفسیر برنامه‌ریزان برگردان می‌شود، سنگاپور (یک ایالت شهری) بعنوان مثال در نظر گرفته می‌شود. با توسعه خط لوله داده‌کاوی، توزیعات فضایی الگوهای سفر را توسط ساکنین مناطق مختلف شهر مقدارسنجی می‌کنیم. هدف نهایی این است که به برنامه‌ریزان کمک شود تا بطور کارآمدی دانش شهری را از داده‌های بزرگ به منظور هدف قرار دادن مناطق شهری مخصوص برای زیرساختار آتی و بهبود برنامه‌ریزی خدمات اشتقاق یا بهره‌برداری کنند.

بقیه مقاله به شرح زیر است. در بخش 2، منابع حالت هنر در کاوش الگوهای تحرک انسانی از داده‌های تلفن همراه را بررسی می‌کنیم. سپس در بخش 3، حوزه مطالعه و داده‌ها را بررسی می‌کنیم که شامل سوابق جزئیات تماس (CDR)، سرشماری، و داده‌های بررسی مسافرت خانوادگی (برای اهداف اعتبارسنجی) است. در بخش 4، روش‌های داده‌کاوی برای استخراج برآوردهای قابل اطمینان از منظر آمار از شبکه‌های تحرک فردی از داده‌های CDR را معرفی می‌کنیم. در بخش 5، اقداماتی را به منظور اندازه‌گیری و تعیین کمیت توزیع فضایی شبکه‌های تحرک در شرایط شهری ارائه می‌کنیم. در نهایت، در بخش 6 درباره پیامدهای برنامه‌ریزی از یافته‌ها برای توسعه آتی شهر بحث می‌کنیم.

2 بررسی منابع

هم چنان که اتصال‌پذیری تلفن همراه مسیر ارتباطات مردم، کار، و بازی را تغییر داده است، از اینرو داده‌های تلفن همراه به منظور دستیابی به اطلاعات فوری مکانی از محل نگهداری ناشناس کاربران تلفن برای آنالیز الگوهای تحرک آنها استفاده می‌شود [14]، [15]، [16]. اگرچه چنین داده‌هایی اغلب در فضا و زمان کمیاب هستند، با اینحال حجم زیاد و دوره مشاهده طولانی از داده‌های تلفن همراه را می‌توان به منظور ردیابی ردپای انسانی در مقیاس بی‌سابقه استفاده کرد [17]، [18]، [19]. بلوندل و همکاران [20] پیشرفت‌های وسیع اخیرا به دست آمده را توسط مطالعات بر روی موضوعات تحرک انسانی، پارتیشن‌بندی جغرافیائی، برنامه‌ریزی شهری، توسعه و امنیت و حفظ حریم خصوصی بررسی می‌کنند. کالابرس و همکاران [21] مطالعه متمرکزی را بر روی ایده‌ها و تکنیک‌ها ارائه می‌کنند که داده‌های تلفن همراه را برای سنجش شهری به کار می‌برد. در این مطالعه، بر استفاده از داده‌های تلفن همراه متمرکز می‌شویم تا تحرک انسانی را درک کنیم. اثر قبلی در این جنبه الگوئی از بازده ترجیحی برای موقعیت‌های بازدید شده قبلی و کاوش اماکن جید بعنوان یک مشخصه کلی و جهانی را نشان می‌دهد [18]، [22]، [23]، [24]. بر اساس این مشخصه، این موضوع محتمل است که موقعیت‌های فعالیت انسانی معنادار با استفاده از داده‌های تلفن همراه برآورد شود. CDRها بعنوان داده‌های قدیمی مسافرت انسانی ایجاد می‌شوند که حاوی اطلاعات مکانی و زمانی برای مقاصد فعالیت معنادار است، یا به بیانی دقیق‌تر داده‌های GPS است که دقت و فرکانس بالاتری را ارائه می‌کند [25]. با اینحال، یک تولید جانبی برای اهداف صورتحساب بطور معمول انجام شده توسط حامل‌های خدمات تلفن همراه، داده‌های CDR

را می‌توان در در هزینه‌ای بسیار کمتر و در مقیاسی بزرگتر به دست آورد. CDRها اطلاعات مکانی و زمانی تحرکات کاربران تلفن همراه را در برج سلولی یا سطح دانه‌ریز، با توجه به فن‌آوری موقعیت‌یابی مکانی به کار رفته توسط حامل‌های خدمات ارائه می‌کنند. در بخش‌های زیر، بطور دقیق مطالعات قبلی را بررسی می‌کنیم که داده‌های CDR را به منظور استنتاج تحرک انسانی در شهرها استفاده می‌کند.

a آنالیز مبتنی بر سفر

وانگ و همکاران [26] روشی را برای تولید ماتریس‌های مبدا-مقصد (OD) گذار مبتنی بر برج برای دوره‌های زمانی مختلف توسعه می‌دهند و آنها را به ODها گذار گره به گره در شبکه جاده‌های برای بوستون و سن فرانسیسکو تبدیل می‌کنند. در زیر یک رویکرد مشابه است که، اقبال و همکاران [27] داده‌های CDR جمع‌آوری شده در داکا پایتخت بنگلادش در طول یک ماه را در ترکیب با داده‌های ترافیکی استفاده می‌کنند، تا ODهای گذار گره به گره را برآورد کنند. در حالی که این روش مسیری را برای تبدیل داده‌های CDR به ODها ارائه می‌کند که برای اهداف برنامه‌ریزی حمل‌ونقل حائز اهمیت است، این به مانند رویکرد مبتنی بر سفر است که بخش‌های سفر از مسافرت مبتنی بر ظاهر مردم در فضا زمان را تقلید یا شبیه‌سازی می‌کند. با اینحال، این می‌تواند مشکل‌ساز باشد زمانی که داده‌های CDR در رزلوشن یا وضوح مکانی پائین (بعنوان مثال، فاصله‌گذاری بین برج‌های سلولی وسیع است، بیشتر از چندین کیلومتر می‌باشد) است اما شبکه‌های جاده‌های در مناطق پوشش برج متراکم است، تخصیص ODها به شبکه جاده‌های مسیره‌های مورب جدا شده‌های را در جاده‌های محلی ایجاد می‌کند.

برای رفع این مسئله، تجزیه خط سیرها به اماکن ثابت که در آنها مردم می‌مانند تا فعالیت‌های خودشان را انجام دهند، حاذق اهمیت است. با توجه به سازگاری کاربردهای تلفن همراه مبتنی بر موقعیت و گوشی‌های هوشمند، مجموعه وسیعی از منابع علم کامپیوتر ظهور کرده است تا موجب پیشرفت و بهبود روش‌هایی به منظور کاوش داده‌های خط سیر شود [28]. هدف یافتن الگوریتم‌های مناسب به منظور استخراج اماکن اقامت معنادار برای آنالیز بیشتر است تا موجب کاهش نویز در داده‌های بزرگ شود. با به کار بردن الگوریتم‌ها برای تجزیه خط سیرهای CDR به موقعیت‌های ایستا، الکساندر و همکاران [29] روش‌هایی را برای برآورد جداول سفر OD با استفاده از زمان روز (صبح، ظهر، عصر،

و بقیه روز) و با هدف سفر، همچون اثرات کاری مبتنی بر خانه (HBW)، سایر آثار مبتنی بر خانه (HBO) از خط سیرهای تلفن همراه سه‌گوش دانه‌ریز برای 2 میلیون کاربر در مدت زمان 2 ماه برای بوستون ارائه می‌کنند. الکساندر و همکاران با استفاده از داده‌های بررسی مسافرت خانوادگی در همان منطقه و تولیدات برنامه‌ریزی حمل‌ونقل سرشماری (CTPP)، یک همبستگی قدرتمندی را در میان ODهای سه منبع در سطح کلانشهری شناسایی می‌کنند، که در عین حال اثربخش استفاده از داده‌های CDR برای هدایت جریان‌های OD در سطح جغرافیائی جمع‌آوری شده را نشان می‌دهد.

b. آنالیز مبتنی بر فعالیت

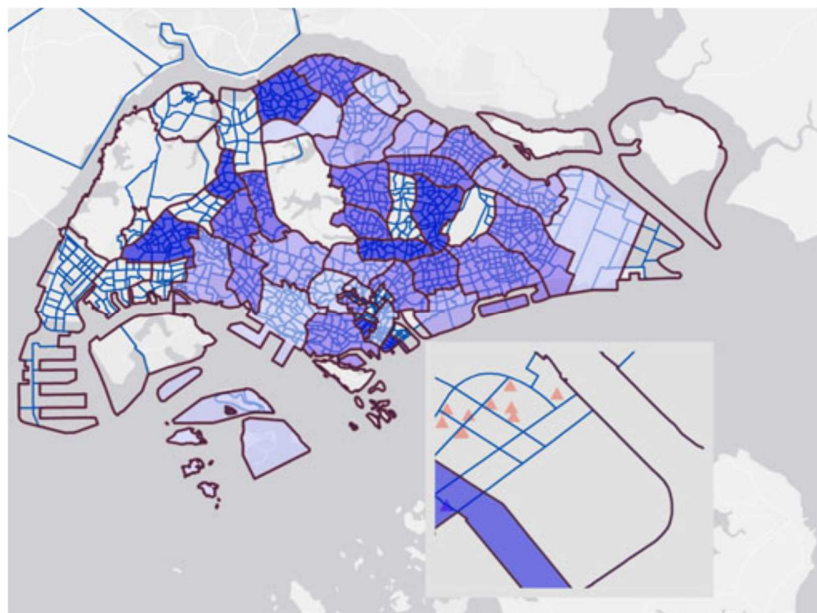
سفرهای روزانه انسانی بر اساس شرایط سفت‌وسختی سازماندهی می‌شود که در زندگی روزمره آنها از اهمیت برخوردار است. از این نقطه نظر، مدل مبتنی بر فعالیت - رویکرد حالت هنر در برنامه‌ریزی حمل‌ونقل - نقضای سفر را بعنوان نیازهای ایجاد شده برای انجام فعالیت‌ها در نظر می‌گیرد [30] - [33]. بنابراین، توسعه روش‌هایی برای برگردان داده‌های بزرگ شهری به یک رویکرد مبتنی بر فعالیت برای مدلسازی شهری و همچنین حمل‌ونقل مرتبط و مهم است. با اتخاذ مفهوم «مایه اصلی یا نقوش» از تئوری شبکه پیچیده [34]، اشنایدر و همکاران [35] شبکه‌های تحرک روزانه از داده‌های CDR را برای شهر پاریس در طول مدت 6 ماه و با استفاده از داده‌های بررسی سفر برای شهرهای پاریس و شیکاگو برای یک یا دو روز بررسی می‌کنند. با استفاده از روش ساده استخراج موقعیت‌های معنادار فعالیتی، آنها برج‌هایی با بازدیدهای مکرر بالاتر از یک آستانه خاص را بعنوان موقعیت‌های بالقوه ایستا (ماندن) حفظ می‌کنند. استدلال اشنایدر و همکاران بر این است که با استفاده از تنها 17 انگاره ویژه، 90 درصد از الگوهای مسافرتی مشاهده شده در هر دو بررسی همراه با مجموعه داده‌های تلفن همراه را می‌توان برای مناطق کلانشهری بازیابی نمود. از طریق تلقی یا قلمداد نمودن بسیار دقیق اماکن اقامت در داده‌های CDR تلفن همراه سه‌گوش دانه‌ریز برای یک میلیون کاربر در بوستون، ژیانگ و همکاران [24] رویکرد مشابهی را برای استخراج نقوش روزانه به کار می‌برند. آنها یافته‌های مشابه را گزارش می‌کنند و روش استنتاجی احتمالی را برای استفاده نقوش، زمان روز، توالی فعالیت، و اطلاعات مربوط به استفاده از زمین برای استنتاج بیشتر انواع فعالیت و تخصیص ترافیک به شبکه‌های حمل‌ونقل مبتنی بر مسافرت

ایجاد شده در این رویکرد را ارائه می‌کنند. ویدهالم و همکاران [36] ایده دستیابی به انواع فعالیت (مانند فعالیت در خانه، کار، خرید، اوقات فراغت، و غیره) برای اماکن اقامتی استخراج شده از داده‌های تلفن همراه و داده‌های استفاده از زمین را برای شهرهای بوستون و وین پیاده‌سازی می‌کنند.

یک نقطه ضعف معمول در این مطالعات فراموشی روش‌های گسترش نمونه به منظور توسعه نتایج مدلسازی شده از کاربران تلفن همراه برای جمعیت حاضر در سطح کلانشهر است. این بطور خاصی مرتبط می‌باشد. از سوی دیگر، ویژگی‌های جمعیت‌شناسی اجتماعی برای کاربران بیشمار تلفن همراه در دسترس نیست. از سوی دیگر، همانطور که مطالعات قبلی برای اهداف برنامه‌ریزی شهری و حمل‌ونقل نشان می‌دهد، روش‌های نمونه‌برداری CDR همانند روش‌های نمونه‌برداری بررسی حائز اهمیت است.

c. رویکرد مبتنی بر سیستم

تول و کلاک و همکاران ([37] و [38]) روش‌های پردازش داده‌های CDR را ترکیب می‌کنند تا تقاضای سطح را برآورد کنند، و چارچوب نوآورانه‌ای را برای هدایت ترافیک برآورد شده در شبکه‌های جاده‌ای ارائه کنند و الگوهای استفاده‌ی جاده از



شکل 1. مناطق برنامه‌ریزی توسعه سنگاپور (DGP)، مناطق آنالیز حمل‌ونقل (MTZ)، و برج‌های سلولی.

داده‌های CDR خام برای شهرها در قاره‌های مختلف را درک کنند. این تلاش‌ها دانش خودمان را درباره چگونگی استفاده از داده‌های CDR در درک تحرک انسانی جهت تولید جداول OD در مقیاس شهری با هزینه‌ای پائین و برآورد زمان‌های مسافرت در شبکه‌های جاده‌های افزایش داده است. همچنین لورنزو و همکاران [39] یک رویکرد مبتنی بر سیستم را نشان می‌دهند. آنها یک ابزار هوشمند به نام AllAboard را توسعه می‌دهند، که داده‌های تلفن همراه را به منظور کمک به مسئولین شهری در کشف و کاوش تحرک شهری و بهینه‌سازی حمل‌ونقل عمومی آنالیز می‌کند. با اینحال، با توجه به پیچیدگی در پیاده‌سازی و اجرا، رویکرد آنالیز غالب به جای روش مبتنی بر سیر یا فعالیت، رویکرد مبتنی بر سفر است. با استفاده از منابع فوق، سنگاپور را بعنوان یک مطالعه موردی استفاده می‌کنیم تا خط لوله پردازش داده‌ها برای استخراج الگوهای تحرک انسانی از داده‌های CDR را توسعه دهیم به این منظور که رفتار مسافرت را درک کنیم و کاربردهای داده‌های بزرگ برای اهداف برنامه‌ریزی شهری و حمل‌ونقل را بهبود بخشیم.

3 حوزه مطالعه و داده‌ها

سنگاپور را بعنوان مطالعه موردی در این مقاله استفاده می‌کنیم. سنگاپور یک دولت‌شهر با محدوده 43 کلومتری در جهت غرب-شرق و 23 کیلومتر در جهت شمال-جنوب است. جمعیت این دولت‌شهر در سال 2010 حدود 5.18 میلیون بود، که از این تعداد 3.79 میلیون ساکن و 1.39 میلیون غیر ساکن یا غیر مقیم بودند. سنگاپور بالاترین نرخ نفوذ تلفن همراه در جهان را دارد، یعنی 150 درصد (با استفاده از کل جمعیت بعنوان مقدار پایه).

a داده‌های سوابق جزئیات تماس

دو هفته متوالی (یعنی 14 روز) از داده‌های سوابق جزئیات تماس تلفن همراه (در ماه‌های مارس و آوریل سال 2011) از یک حامل را استفاده می‌کنیم تا الگوهای تحرک افراد ناشناس در منطقه کلانشهری را بررسی کنیم. مجموعه داده‌های دوره مطالعه شده حاوی 3.17 میلیون کاربر ناشناس تلفن همراه، و جمعاً 722.92 سوابق استفاده از تلفن همراه است. بیش از 5 هزار برج تلفن همراه در سنگاپور، با فاصله‌گذاری 50 متری در مناطق متراکم قسمت مرکزی شهر تا چند کیلومتر در مناطق حومه وجود دارد. در کل، شبکه برج تلفن همراه تراکم بسیار بالایی را دارد که کل منطقه کلانشهر را تحت پوشش قرار می‌دهد (شکل 1 را ببینید).

b. داده‌های سرشماری و نواحی جغرافیائی

علی‌رغم نرخ بالائی از نفوذ، داده‌های به دست آمده از یک حامل تلفن همراه برای این مطالعه تنها 63 درصد از کل جمعیت سنگاپور را شامل می‌شود. از نقطه نظر آماری برای دستیابی به مقیاس‌های معنادار از داده‌های CDR برای کل جمعیت، نیاز داریم تا کاربران تلفن همراه نمونه‌برداری شده را گسترش دهیم. بنابراین داده‌های سرشماری با اطلاعات مکانی مفید است. سرشماری عمومی موجود در سنگاپور شامل جمعیتی توسط گروه‌های مختلف جغرافیائی در سطح ناحیه برنامه‌ریزی (DGP یا همان منطقه برنامه‌ریزی راهنمای توسعه) است. فرض را بر این می‌گیریم که یک فرد بالای 10 سال ممکن است دارای تلفن همراه باشد، کاربران در داده‌های CDR به جمعیت ساکنین سنگاپور در این رده سنی در سطح DGP گسترش یابد.

DGP یک واحد مکانی از منطقه برنامه‌ریزی استفاده شده توسط آژانس برنامه‌ریزی شهری سنگاپور - اداره بازسازی شهری (URA) است. 56 مورد DGP در سنگاپور با طیف‌بندی اندازه‌ها از 0.85 sq km تا 66 sq km وجود دارد. 35 تا از 56 مورد DGP دارای جمعیت مسکونی ساکن در آنها هستند، و بقیه یا در مناطق صنعتی یا در زمین‌های متعلق به سایر اهداف ساکن هستند. نواحی DGP بیشتر به مناطق مکانی یا فضایی ریزدانه زیر تقسیم می‌شوند، نواحی آنالیز حمل‌ونقل، که به اصطلاح MTZها هستند، برای اهداف برنامه‌ریزی شهری توسط سازمان برنامه‌ریزی - حمل‌ونقل سنگاپور، اداره حمل‌ونقل و زمین (LTA) استفاده می‌شود. با اینحال، داده‌های جمعیتی در این سطح از وضوح مکانی بالا (یعنی MTZ) در دسترس عموم نیست. در مجموع حدود 1100 از MTZ کل منطقه شهری را تحت پوشش قرار می‌دهد، و طیف‌بندی مقادیر آنها از 0.015 sq km تا 43 sq km است.

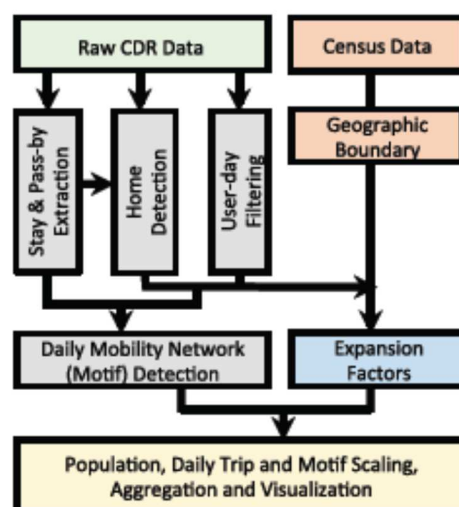
شکل 1 تراکم جمعیت را در سطح DGP نشان می‌دهد، و رابطه مکانی DGP، MTZ، و برج‌های سلولی (تلفن همراه) را به نمایش می‌گذارد. کاربران تلفن همراه را در سطح برج به منظور انطباق کل جمعیت در سطح DGP گسترش می‌دهیم، و جمعیت را در سطح MTZ تخمین می‌زنیم. درباره جزئیات گسترش در بخش 4 بحث می‌کنیم.

c. داده‌های بررسی

از مصاحبه درباره داده‌های بررسی مسافرت خانوادگی سنگاپور (HITS) 2008 جمع‌آوری شده توسط LTA سنگاپور استفاده می‌کنیم تا نتایج برآورد شده از داده‌های CDR را اعتبارسنجی کنیم. در مقایسه با CDRها، داده‌های بررسی سفر اغلب در در مقادیر نمونه کوچک است و با عوامل گسترش برای اهداف توسعه نمونه‌های بررسی برای کل جمعیت همراه است. عوامل گسترش بررسی از طراحی نمونه‌برداری مبتنی بر جمعیت‌شناسی افراد (بعنوان مثال سن، جنسیت و غیره) است. داده‌های بررسی شامل اطلاعات دقیقی درباره آمار خانوادگی و اجتماعی فرد و سوابق خودگزارشی سفر توسط پاسخ دهندگان به بررسی یا نظرسنجی می‌باشد. HITS سنگاپور 2008 شامل 34000 نفر و اطلاعات سفر یک روزه آنها (بعنوان مثال زمان سفر و زمان خروج و اتمام سفر، مقصد و محل‌های مسافرت و غیره) است. در این مطالعه، زمان مسافرت و اطلاعات موقعیت و عوامل گسترش را استفاده می‌کنیم تا نمونه‌های نظرسنجی را به کل جمعیت تعمیم و بسط دهیم، و نتایج حاصله را با تخمین‌هایی از داده‌های CDR مقایسه می‌کنیم.

4. روش‌ها

برای درک الگوهای تحرک انسان در سطح کلانشهر برای اهداف برنامه‌ریزی شهری و حمل‌ونقل، در این مطالعه، روش‌ها در تحقیق قبلی را ترکیب می‌کنیم [24]، [29]، [35]، و خط لوله‌ای را توسعه می‌دهیم از (1) تجزیه داده‌های CDR به منظور استخراج اماکن اقامتی کاربران تلفن؛ (2) شناسایی موقعیت خانگی کاربران تلفن؛ (3) فیلتر کردن کاربران و انتخاب آماری نمونه‌های انگاره‌ای از داده‌های CDR تجزیه شده؛ (4) شناسایی شبکه‌های



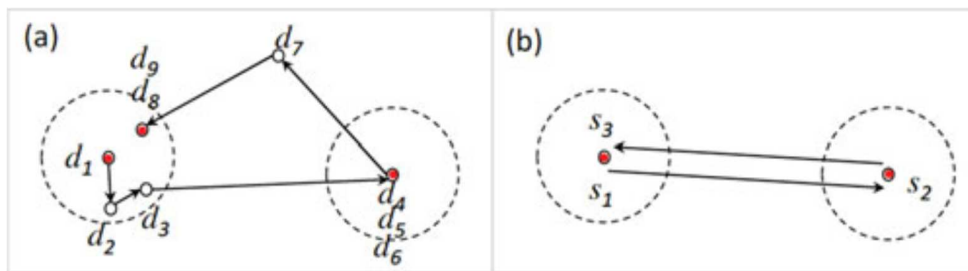
شکل 2. گردش کاری برآورد جمعیت و استخراج الگوی تحرک از داده‌های CDR.

تحرك روزانه از کاربران تلفن؛ (5) جمع‌آوری عوامل گسترش برای مشاهدات فیلتر شده تلفن توسط تركيب كردن داده‌های تلفن پردازش شده و داده‌های آمار؛ گسترش کاربران تلفن همراه، سفرها، و نقوش روزانه؛ و جمع‌آوری آنها از برج برای نواحی آنالیز حملونقل (شکل 2 را ببینید).

a تجزیه خط سیر برای استخراج اماکن

تجزیه داده‌های CDR تلفن همراه به منظور استخراج نقاط anchor (یعنی اماکن توقف) از مسافرت روزانه فردی و تمایز توقف‌ها از نقاط عبور نخستین مرحله قبل از شناسایی شبکه‌های تحرك فردی است. ژنگ [28] مرور جامعی را در داده‌کاوی خط سیر ارائه می‌کند. با الهام از الگوریتم معرفی شده توسط هاریهاران و توپاما [40] که قبلاً با استفاده از داده‌های تلفن همراه سه‌گوش تست و بررسی شده است [24]، [29]، الگوریتم برای خط سیرهای CDR مبتنی بر برج را در این مطالعه اصلاح می‌کنیم. سایر الگوریتم‌ها در تشخیص ایستادن در منابع یافت می‌شود [41]، [42]. در حالی که تشخیص توقف می‌کند، قصد ما دستیابی به سه هدف است: (1) حذف دو نوع از نویزها در داده‌های CDR خام - داده‌های خارج از محدوده و پرش‌های سیگنال بین برج‌های سوابق موقعیت مکانی جغرافیایی کاربران - تا موقعیت واقعی آنها شناسایی شود، (2) خوشه‌بندی نقاطی که از منظر مکانی (در آستانه Δd) نزدیک است و بطور موقتی (در توالی) مجاور به یک موقعیت مستقل است، همراه با یک مدت زمان خرد شده از این خوشه‌بندی بعنوان محل اقامت یا ماندن، و (3) نقاط متراکمی که از نقطه نظر مکانی نزدیک است اما لزوماً مجاور به بعد زمانی در یک مکان منحصر بفرّد نیست.

- با در نظر گرفتن یک موقعیت i کاربر ثبت شده در CDR در توالی بعنوان $D_i = (d_i(1), d_i(2), \dots, d_i(n_i))$ ، با انتخاب فاصله رومینگ Δd_i (بعنوان مثال 300 متر) بعنوان آستانه، از منظر اکتشافی موقعیت‌های نزدیک مکانی را در Δd_i برای هر نقطه در مدوید آنها (نقطه در یک مجموعه که فاصله حداکثری برای سایر نقاط در مجموعه را کمینه می‌کند) خوشه‌بندی می‌کنیم و یک توالی جدید $D'_i = (d'_i(1), d'_i(2), \dots, d'_i(n'_i))$ را تشکیل می‌دهیم، که در آن $(tower(k), t(k)) = d_i(k) =$ برج (k) و $t(k)$ آیدی برج است، و نشان زمانی مشاهده k ام کاربر i در داده‌های خام CDR است. $d'_i(g) = (cellid(g), t(g), dur(g))$ شامل آیدی برج است، که از زمان



شکل 3. شناسایی نقطه ماندن (محل اقامت).

برج مشاهده شده نخست در خوشه شروع می‌شود، و مسافتی است که کاربر در خوشه می‌ماند.

- سپس نقاط در مجموعه توالی خط سیر D_i مبتنی بر آستانه مسافت Δd_2 (بعنوان مثال 300 متر) خوشه‌بندی می‌شود، و نقاط حفظ می‌شود که فاصله آنها بیشتر از آستانه زمانی Δt (برای مثال 10 دقیقه) است بعنوان نقاط نهایی ماندن $S_i = (s_i(1), s_i(2), \dots, s_i(m_i))$ ، که در آن $s_i(m) = (cellid(m), t(m), dur(m))$. پس با انجام این کار، داده‌های خارج از محدوده در فرایند خوشه‌بندی را حذف می‌کنیم. شکل 3 ورودی و خروجی فرایند را نشان می‌دهد.

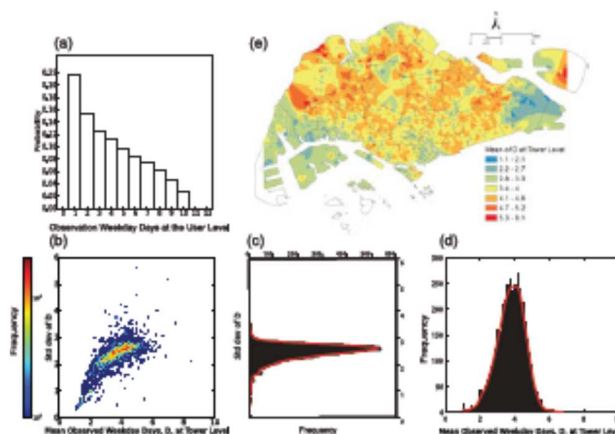
b. شناسایی خانه

این موضوع مهم است که بنا به چندین دلیل محل اقامت کاربران تلفن همراه (برج) علامت‌گذاری شود. نخست، زمانی که بعداً کاربران را به جمعیت بسط دهیم، نیاز است که داده‌های CDR را با داده‌های سرشماری ترکیب کنیم، که تنها جمعیت را توسط موقعیت مسکونی ثبت می‌کند. دوم، به منظور شکل‌دهی شبکه‌های تحرک، فرضیه‌ای را ایجاد می‌کنیم که برای هر روز، یک فرد همیشه از خانه خارج می‌شود و تا انتهای روز به خانه برمی‌گردد. اگرچه در جهان واقعی چندین استثناء وجود دارد، این فرضیه مسیر را ساده‌سازی می‌کند تا شبکه‌های تحرک را از داده‌های CDR استخراج کنیم و این اجازه را به ما می‌دهد که گردش‌ها را با روش توجیه‌پذیری برآورد کنیم. از آنجا که داده‌های CDR بطور منفعلانه اطلاعات فضایی و زمانی کاربران تلفن همراه را ثبت می‌کند، پس همیشه اطلاعات کاملی را درباره اماکن تقریبی افراد نمی‌دهد. سوم، به دلیل هدف برنامه‌ریزی، این امر مهم است که الگوهای تحرک افراد از خانه آنها درک شود، همچون محیط ایجاد شده و اسفاده از زمین محل اقامت فردی که فعالیت‌ها و مسافرت آنها را تحت تاثیر قرار می‌دهد [43]، [44]. قادر بودن به شناسایی الگوهای تحرک انسانی مربوط به فضای شهری برنامه‌ریزان را قادر می‌سازد تا بهبود و توسعه آتی شهری را هدف قرار دهند.

برج تلفن همراه کاربر بعنوان برج ارتباطی غالب در طول شب‌های ایام هفته، و آخر هفته در طول دوره مطالعه (یعنی 14 روز در این مطالعه) شناسایی می‌شود. تعریف «شب» پارامتری است که در زمینه‌های مختلف شهری تنظیم می‌شود. یک شب را از 7 عصر تا 7 صبح در این مطالعه بر اساس شرایط محلی در سنگاپور تعریف می‌کنیم. پس از پردازش داده‌های CDR، 2.88 میلیون کاربر به دست می‌آوریم که برج‌های خانه آنها را می‌توان در داده‌های 2 هفته‌ای مشاهده نمود. این شامل 91 درصد از کاربران تلفن همراه موجود در داده‌های برای دوره مطالعه است، و حدود 56 درصد از کل جمعیت می‌باشد. با توسعه مناسب این کاربران، می‌توان جمعیت را از سطح منطقه برنامه‌ریزی متراکم (DGP)، تا سطح منطقه آنالیز حمل‌ونقل محله (MTZ) با وضوح مکانی بالا تجزیه نمود. روش گسترش جمعیت و نتایج را در بخش 4.5 بحث می‌کنیم.

c. فیلترینگ نمونه‌های روزانه کاربر تلفن همراه

در مقایسه با داده‌های بررسی قدیمی، مزیت داده‌های CDR دوره طولانی‌تری از مشاهده



شکل 4. آمارهای روز کاربر پس از فیلترینگ نمونه: (الف) توزیع فراوانی تعداد کلی ایام هر کاربر؛ (ب) توزیع تراکم انحراف استاندارد (نشان داده شده در ج) و میانگین (نشان داده شده در د) تعداد کاربران از روزهای مشاهده در سطح برج؛ (و) توزیع مکانی میانگین تعداد کاربران از روزهای مشاهده در سطح برج.

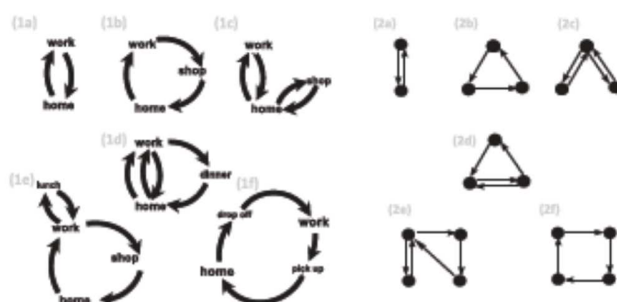
و اندازه بزرگتر نمونه است، و نقطه ضعف آن کم پستی یا تنگی داده‌ها است. CDRها ممکن نیست مسافرت فردی را بعنوان خاطرات سفر یا کاربردهای ردیابی که Gps انجام می‌دهد، آشکار سازد. پس این موضوع مهم است که روزهای کاربر نمونه‌برداری شود زمانی که دستگاه تلفن همراه به حد کافی مورد استفاده قرار می‌گیرد. تحقیق قبلی [35] به

این نتیجه رسیده است که اگر فعالیت‌های ارتباطی تلفن همراه کابر متجاوز از آستانه خاصی باشد، داده‌های CDR را می‌توان بطور مشابهی بعنوان داده‌های بررسی تلقی کرد، و از نقطه نظر آماری نتایج سازگار و قیاس‌پذیری را با توجه به الگوهای تحرک فردی ارائه می‌کند. در این مطالعه، قوانین مشابهی را برای فیلترینگ نمونه به شرح زیر اتخاذ می‌کنیم.

- تنها یک مشاهده از روز کاربر را حفظ می‌کند، اگر که در یک روز (یعنی 24 ساعت) کاربر دارای سوابق تلفن در حداقل هشت ساعت از اسلات‌های زمانی مجزا از 48 اسلات زمانی نیم ساعته باشد.

- جداسازی مشاهدات در ایام هفته از کشاهداتی که در آخر هفته هستند، هم چنان که الگوهای تحرک متفاوت هستند. در اینجا فقط روس سوابق در آخر هفته متمرکز می‌شویم.

پس از فیلترینگ، 6.28 میلیون ایام کاربر برای 1.55 میلیون کاربر در 10 روز هفته از دو هفته را حفظ می‌کنیم. تعداد کلی کاربران پس از فیلترینگ 49 درصد از تمامی کاربران در داده‌های خام است. توزیع مقدار ایام کاربر در هر کاربر پس از فیلترینگ در شکل 4الف را ترسیم می‌کنیم. می‌توانیم ببینیم که، برای این 1.55 میلیون کاربر فیلتر شده، 22 درصد از آنها تنها دارای 1 مشاهده روز هستند، 11 درصد دارای 4 روز، 7 درصد دارای 7 روز، و سه درصد دارای 10 روز هستند. بطور متوسط، هر کاربر دارای 4 مشاهده روزانه است. همچنین اطلاعات کاربر برای سطح برج را جمع‌آوری می‌کنیم. شکل 4ب توزیع تراکم انحراف استاندارد (نشان داده شده بصورت مجزا در شکل 4ج) از تعداد کاربران ایام مشاهده در سطح برج مبتنی بر برج خانه کاربران را نشان می‌دهد. توزیع مکانی متوسط تعداد کاربران ایام مشاهده در سطح برج را در شکل 4د ترسیم می‌کنیم. می‌توانیم ببینیم که در مرکز شهر، فرودگاه، و پایان غربی منطقه ساحلی، تعداد روز کاربر بر هر کاربر به نسبت کم است. احتمالاً، گردشگران و خارجی‌ها در این مناطق ساکن هستند. در حالی که در نواحی مسکونی در حومه‌های شهر، همچون آنگ‌موکیو، باکیت باتوک، ژورانگ وست، پانگل، سنگانگ، تامپینس، پیشون، و وودلندز، متوسط تعداد ایام کاربر نسبت به مقدار میانگین نسبتاً بیشتر است.



شکل 5. مثال‌هایی از شبکه‌های تحرک روزانه (الف-1و) در بررسی‌های سفر روزانه و (الف-2و) در انتزاع.

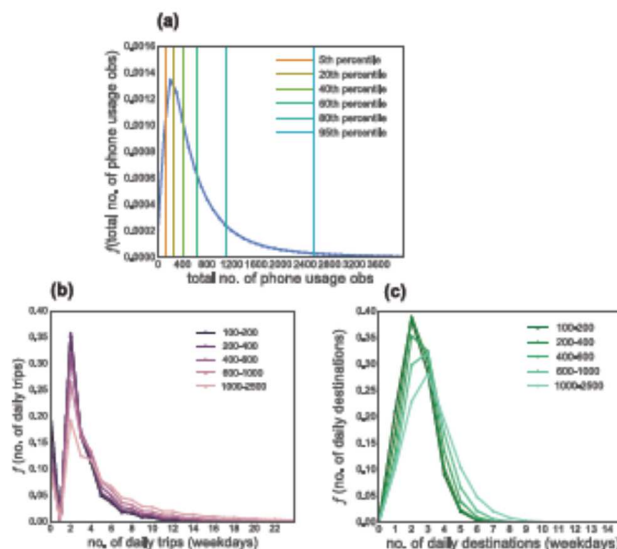
4.4. شناسایی الگوهای تحرک مبتنی بر فعالیت

4.4.1. نقوش تحرک روزانه

تحرک روزانه انسان را می‌توان توسط چندین فعالیت ضروری برای زندگی به شدت ساختاری سازماندهی نمود. شناسایی شبکه‌های تحرک فردی (یعنی نقوش یا موتیف‌های روزانه) حائز اهمیت است. مدل‌های تقاضای مسافرتی معاصر توسعه یافته و استفاده شده توسط سازمان‌های حمل‌ونقل منطقه‌ای در کشورهای در حال توسعه مبتنی بر فعالیت است [33]. شکل 5 (الف-1و) برخی از مثال‌ها درباره الگوهای فعالیت فردی (سفرهای روزانه) معمولاً مشاهده شده در بررسی‌های سفر را نشان می‌دهد. همچنان که در زبان مدل‌سازی حمل‌ونقل اشاره شده است، آنها عبارتند از: (الف) یک سیر کاری مبتنی بر خانه، (ب) یک سیر کار مبتنی بر خانه با یک مقصد سوم، و کار بعنوان مقصد اصلی، (ج) یک سیر کاری مبتنی بر خانه با یک زیر سیر مبتنی بر خانه، (د) دو سیر کار مبتنی بر خانه با یک مقصد سوم، (و) یک سیر کار مبتنی بر خانه با یک زیر سیر مبتنی بر کار، و (ی) یک سیر کار مبتنی بر خانه با یک سیر اسکورت (برای ها کردن و برداشتن برخی از افراد). این الگوهای فعالیتی وارد ترکیبات متعددی می‌شود. با اینحال، اگر یک نمودار بسیار انتزاعی را به منظور ارائه این زنجیره‌های فعالیتی استفاده کنیم، آنها به نقوش روزانه کسب شده توسط داده‌های تلفن همراه پسیو و منفعل کاهش می‌یابند. شکل 5 (الف-2و) (الف-1و) را در فرمت انتزاعی تری نشان می‌دهد، که موتیف‌ها یا نقوش روزانه نامیده می‌شود. آنها بعنوان مقیاس‌های تحرک انسان در مطالعات قبلی برای شهرهای پاریس، شیکاگو و بوستون ارائه و تست و بررسی شده است [24]، [35]، و استدلال شده است که با بررسی‌های سفر در آمارهای جمع‌آوری شده قیاس‌پذیر است.

در اینجا الگوریتمی را به منظور شناسایی نقوش روزانه برای کاربران فیلتر شده تلفن همراه ارائه می‌کنیم که سوابق تلفن همراه آنها در یک روز نمونه‌برداری شده به اندازه کافی (توضیح داده شده در بخش 4.3) برای ارائه الگوی فعالیت روزانه آنها تکرار شده است. به بیان ساده و رسمی، موتیف یک کلاس معادلی از نمودارهای هدایت شده یا مستقیم است. یک نمودار مستقیم یک جفت سفارشی $D = (V, A)$ است که در آن V مجموعه‌ای از گره‌ها (یعنی لبه‌های مستقیم) است، و A یک مجموعه از جفت‌های سفارشی گره‌ها (یعنی لبه‌های مستقیم) است. بدون از دست دادن و تلفات کلیتی، زمانی که V کی مجموعه از عناصر n است، V را این گونه به دست می‌آوریم که مجموعه $V_n \triangleq \{1, 2, \dots, n\}$ است. فرض کنید \mathcal{D}_n مجموعه نمودارهای مستقیم با گره‌های n باشد، یعنی $\mathcal{D}_n \triangleq \{D_n \mid D_n = (V_n, A), A \subseteq V_n \times V_n\}$. نقوش $[D_n]$ برچسب‌گذاری یا نشان‌گذاری گره‌ها را نادیده می‌گیرد.

با امکان اقامت استخراج شده، از جمله محل‌های اقامت بالقوه و مشاهده شده (تشریح شده در بخش 4.1) و محل‌های اقامت شناسایی شده کاربر (شرح داده شده در بخش 4)، قادر به شناسایی شبکه‌های تحرک هستیم همانطور که در شکل 5 (الف-2) نشان داده شده است. در می‌یابیم که نه تنها تمامی کاربران لزوماً دارای نخستین محل اقامت (یا نقطه اقامت بالقوه) هستند که از برج خانه آنها شروع می‌شود، یا آخرین نقطه اقامت یا ماندن (یا نقطه اقامت بالقوه) را دارند که هر روز در برج خانه آنها در خط سیر فیلتر شده به انتها می‌رسد، البته



شکل 6. توزیع فراوانی (الف) تعداد کلی استفاده از تلفن همراه برای کاربران فیلتر شده، (ب) تعداد مسافرت‌های روزانه برای نمونه‌های فیلتر شده روز کاربر توسط گروه کاربر، و (ج) تعداد روزانه از موقعیت‌های منحصریفرده برای نمونه‌های فیلتر شده روز کاربر توسط گروه کاربر.

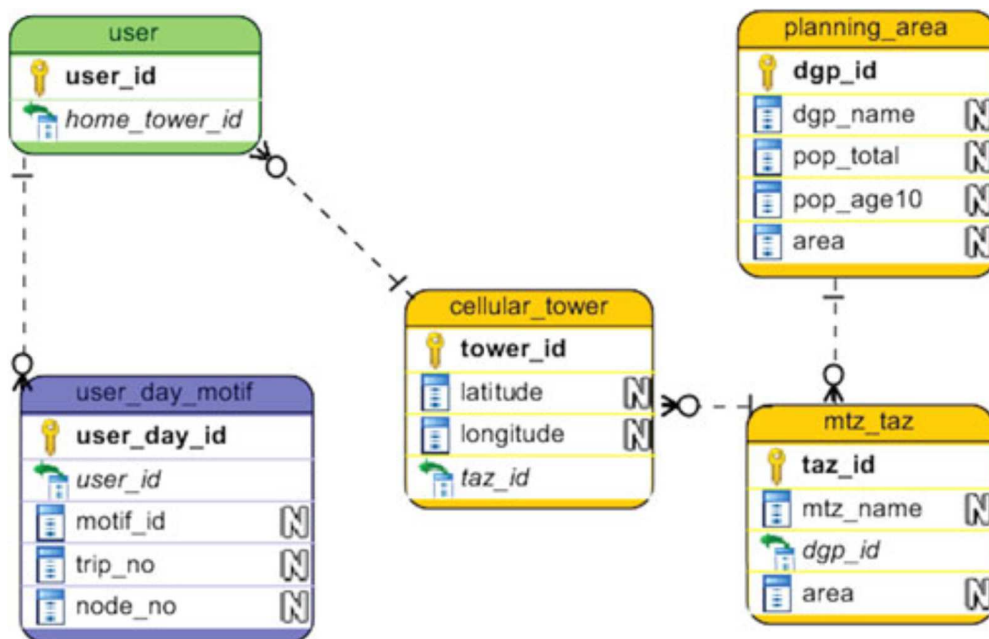
لازم به ذکر است که این موارد با توجه به ماهیت منفعلانه از داده‌های CDR می‌باشد. به منظور شکل‌دهی یک سیر یا تور کامل برای هر کاربر، فرضیه‌ای را ایجاد می‌کنیم که، اگر کاربری در یک روز مسافرت کند، وی گردش نخست را از خانه شروع و گردش آخر را در خانه به اتمام می‌رساند.

- با توجه به نقاط اقامت استخراج شده و روز(های) فیلتر شده، توالی مقاصد در روز نمونه‌برداری شده I_i برای یک کاربر ناشناس مشخص I_i را به دست می‌آوریم، که توسط $S_{ik} = (s_{ik}(1), s_{ik}(2), \dots, s_{ik}(m_{ik}))$ نشان داده شده است.
- خانه h_i کاربر I_i در مرحله قبلی شناسایی می‌شود.
- همیشه بررسی کنید اگر برای روز k ، موقعیت‌های فعالیت اول و آخر کاربر I_i عبارت از h_i باشد. در غیر اینصورت، خانه h_i را به توالی محل اقامت یا همان نقطه ماندن اضافه می‌کنیم، که یک توالی جدید $S'_{ik} = (s'_{ik}(1), s'_{ik}(2), \dots, s'_{ik}(m'_{ik}))$ را تشکیل می‌دهد.
- تعداد کل سفرها (که مسافرت‌ها از مبدا به مقاصد است)، h_{ik} ، برای کاربر I_i در روز k را محاسبه کنید.
- کل تعداد گره‌های متمایز، n_{ik} ، از خط سیر S'_{ik} را محاسبه کنید.
- از یک شبکه (نمودار مستقیم) D_{ik} با گره‌های n_{ik} : با استفاده از یک ماتریس $n_{ik} \times n_{ik}$ یعنی M_{ik} ، شبکه تحرک برای کاربر I_i در روز k را نشان می‌دهیم. اگر حداقل در یک سفر از گره o به گره d باشد، $M_{ik}(o, d) = 1$ ؛ بعبارتی دیگر $M_{ik}(o, d) = 0$.
- موتیف یعنی کلاس معادل D_{ik} را به دست آورید.

4.4.2. الگوهای استفاده از تلفن همراه و الگوهای تحرک

پس از فیلترینگ نمونه‌های روز کاربر، این موضوع مهم است که کاربران با رخدادهای استفاده از تلفن همراه بررسی شوند که دارای تفاوت‌های نظاممند در رفتار مسافرت نیست. بدین منظور، رابطه بین الگوهای استفاده از تلفن همراه

کاربران فیلتر شده و الگوهای سفر روزانه آنها را بررسی می‌کنیم. شکل 6 الف توزیع متناوب از تعداد کلی مشاهدات استفاده از تلفن همراه برای تمامی 1.55 میلیون کاربر فیلتر شده در 10 روز را نشان می‌دهد. پنجمین، 20، 40، 60، 80، و 95امین درصد از کل استفاده از تلفن در طول این دوره برای تمامی کاربران فیلتر شده به ترتیب 140، 262، 420، 650، 1108، و 2558 است. بر اساس



شکل 7. نمودار رابطه نهادی کاربر تلفن همراه پردازش شده، داده‌های تحرک روزانه کاربر از CDRهای آنها، و واحدهای مکانی جغرافیایی: برج‌ها، نواحی آنالیز حمل‌ونقل (MTZها)، منطقه برنامه‌ریزی (DGPها) با جمعیت سرشماری.

این اطلاعات، کاربران فیلتر شده را در 5 گروه، با کل مشاهدات استفاده از تلفن همراه از $(100, 200)$ ، $(200, 400)$ ، $(400, 600)$ ، $(600, 1000)$ ، و $(1000, 2500)$ را گروه‌بندی می‌کنیم. الگوهای سفر روزانه کاربران فیلتر شده از جمله تعداد روزانه مسافرت‌ها، و تعداد روزانه مقاصد منحصر بفرورد برای هر یک از گروه‌ها را مورد بررسی قرار می‌دهیم. شکل 6 ب نشان می‌دهد که برای این 5 گروه از کاربران فیلتر شده، توزیعات فراوانی تعداد روزانه آنها از مسافرت‌ها کاملاً مشابه است. بطور مشابه، شکل 6 ج نشان می‌دهد که توزیعات فراوانی تعداد روزانه اماکن منحصر بفرورد بازدید شده توسط کاربران فیلتر شده‌ی گروه‌های مختلف الگوهای مشابهی را دنبال می‌کنند.

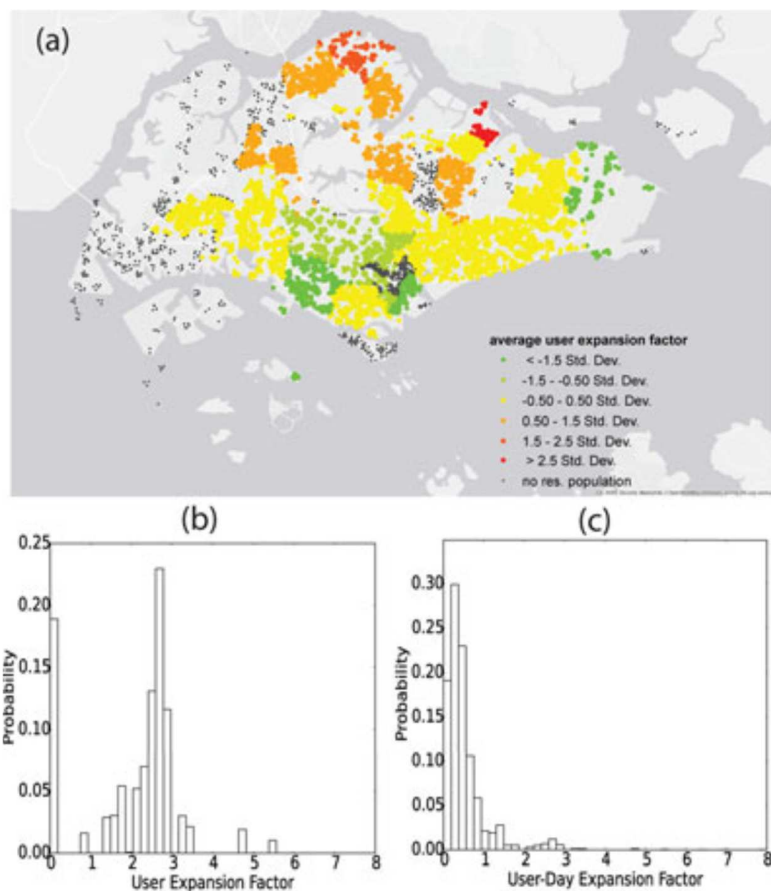
این نشان می‌دهد که کاربران تلفن همراه با الگوهای مختلف استفاده از تلفن دارای تفاوت‌های نظام‌مند در رفتار مسافرت نیستند. این تأیید برای اعتبارسنجی استفاده از نمونه‌های فعال روز کاربر تلفن همراه به منظور بررسی بیشتر الگوهای تحرک روزانه انسان مهم است، و این نمونه‌ها برای گسترش به جمعیت استفاده می‌شود.

4.5. گسترش نمونه‌های تلفن همراه به جمعیت

برای دستیابی به برآوردهای الگوهای تحرک روزانه (نقوش) و سفرها، و الگوهای مکانی قوی برای جمعیت در منطقه کلانشهر، و اطلاعات برای برنامه‌ریزی شهری و حمل‌ونقل، نیاز داریم تا نمونه‌های تلفن همراه (از جمله کاربران و ایام کاربر) را به درستی گسترش دهیم. در این بخش، فرایند را در تولید عوامل گسترش برای داده‌های CDR تشریح می‌کنیم.

4.5.1 عوامل گسترش

دو نوع از عوامل گسترش ایجاد می‌شود که عبارتند از: (1) عوامل گسترش کاربر، (2) عوامل گسترش روز کاربر. شکل 7 نمودار رابطه نهادی مفید در تولید این عوامل گسترش را نشان می‌دهد. پس از شناسایی برج‌های خانه کاربران تلفن همراه، اطلاعات کاربران را در یک جدول پایگاه داده «کاربر»، با شمار id ناشناس آنها ذخیره می‌کنیم، لازم به ذکر است که شماره هر کاربر منحصر بفرد است. الگوهای تحرک کاربران فیلتر شده را که دارای فعالیت‌های ارتباطی تلفن همراه (در حداقل 8 اسلات زمانی نیم‌ساعته در یک روز) در جدول «موتیف روز کاربر» هستند ذخیره می‌کنیم، که



شکل 8. توزیع مکانی عوامل گسترش کاربر در سطح برج، و توزیع فراوانی (ب) عوامل گسترش کاربر، و (ج) عوامل گسترش روز کاربر.

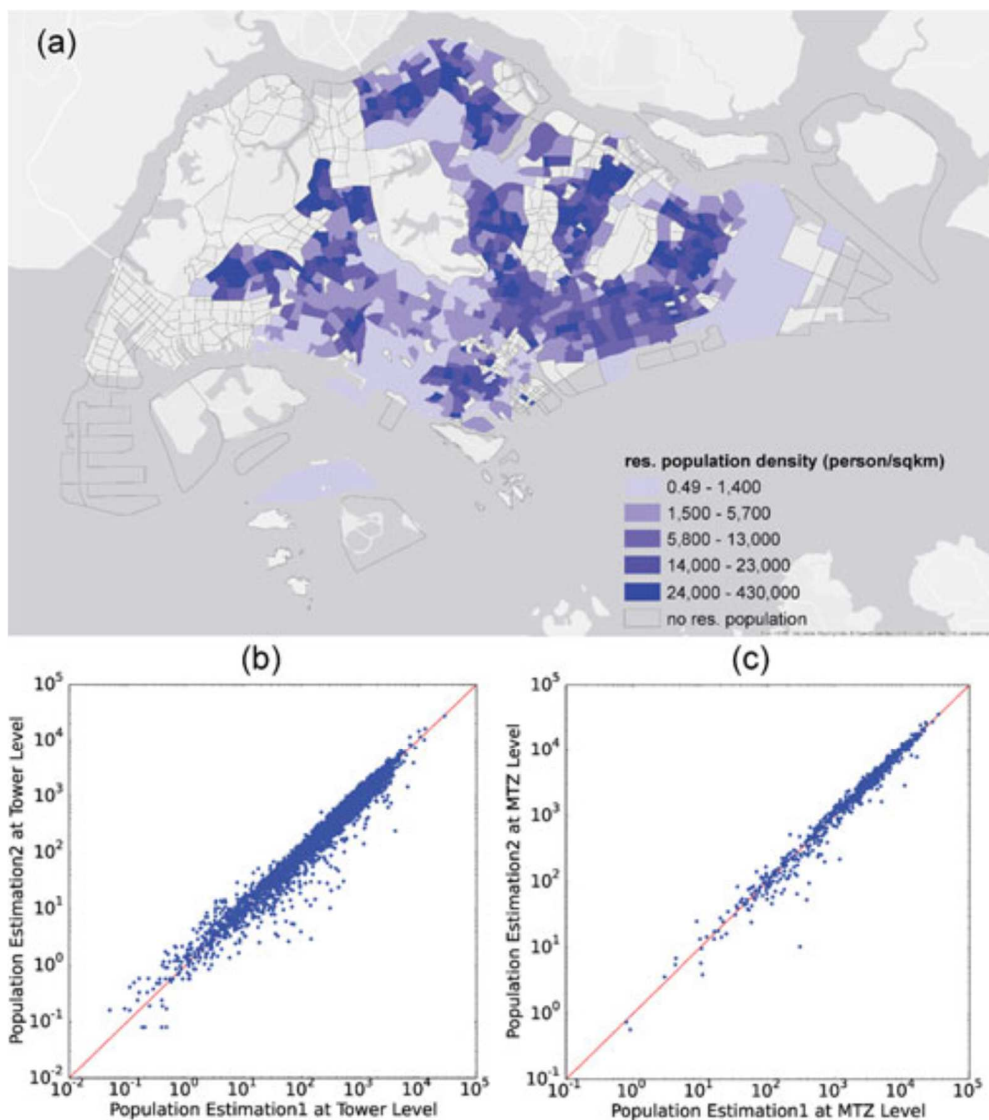
در آن هر ردیف مشاهده مخصوصی را از روز کاربر با id کاربر، id موتیف، و تعداد اطلاعات سفر ارائه می‌کند. 6.28 میلیون سابقه یا پرونده روز کاربر در این جدول برای 1.55 میلیون کاربر منحصر بفرد موجود است که اطلاعات برج خانه آنها را می‌توان در جدول «کاربر» دید. بقیه 1.33 میلیون کاربر، که در جدول «کاربر» هستند اما در جدول «موتیف روز کاربر» وجود ندارند، مواردی هستند که فعالیت‌های ارتباطی تلفن همراه آنها به اندازه‌ای نیست که بتواند آستانه فیلترینگ را برآورده سازد.

همانطور که در شکل 7 نشان داده می‌شود، جدول «کاربر» به جدول «برج سلولی» از طریق «آی‌دی برج سلولی» «آی‌دی برج خانه» کاربران متصل می‌شود. با عملیات مکانی اتصال برج‌های سلولی به مناطق MTZ که آنها می‌افتند، هر برج سلولی مربوط به یک MTZ است. برای این برج‌ها، که در هر MTZ موجود نیست، نزدیک‌ترین MTZ در

فاصله 200 متری استفاده می‌شود تا برای خطای اندازه‌گیری در نمایه مکانی موقعیت برج و منطقه مجاز باشد. جدول «برج سلولی» متصل به جدول «mtz-taz» (از طریق «mtz-id») است، که به جدول «منطقه برنامه‌ریزی» (از طریق «dgp-id») متصل است، که شامل 56 منطقه DGP است که شامل 35 نفر ساکن است.

دو مجموعه از عوامل گسترش در نظر گرفته می‌شود. یک مجموعه شامل عوامل گسترش کاربر برای 2.88 میلیون کاربر با موقعیت خانه قابل شناسایی است؛ مجموعه دیگر شامل عوامل گسترش برای 1.5 میلیون کاربر و عوامل گسترش روز کاربر برای 6.28 میلیون مشاهده از روز کاربر (از 1.55 میلیون کاربر) است. برای مورد اول، عامل گسترش کاربر برای هر کاربر با برج خانه در همان منطقه برنامه‌ریزی مشابه است. برای مورد بعدی، هم چنان که هر کاربری ممکن است دارای مشاهداتی برای بیش از یک روز باشد، عامل گسترش را توسط تعداد کلی ایام مشاهده برای این کاربر عادی‌سازی می‌کنیم. عامل گسترش کاربر β_i را برای کاربر i تعریف می‌کنیم، و عامل گسترش روز کاربر θ_{ik} را برای کاربر i در روز k به شرح زیر است.

- $\beta_i = P_g / \sum_1^{Q_g} U_{h_g}$ ، برای کاربر i با یک برج خانه h_g ، که در آن P_g جمعیت (بالای 10 سال) در منطقه g از DGP گزارش شده در سرشماری است؛ U_{h_g} تعداد کلی کاربران است که برج خانه آنها در g منطقه DGP است، و Q_g تعداد برجها در ناحیه g است؛



شکل 9. (الف) برآوردهای تراکم جمعیت مسکونی؛ (ب) مقایسه برآورد جمعیت از نمونه با تمامی کاربران با خانه شناسایی شده (a.k.a برآورد 1)، و کاربران فیلتر شده با شبکه‌های تحرک کامل (a.k.a برآورد 2) در سطح برج، و (ج) در سطح MTZ.

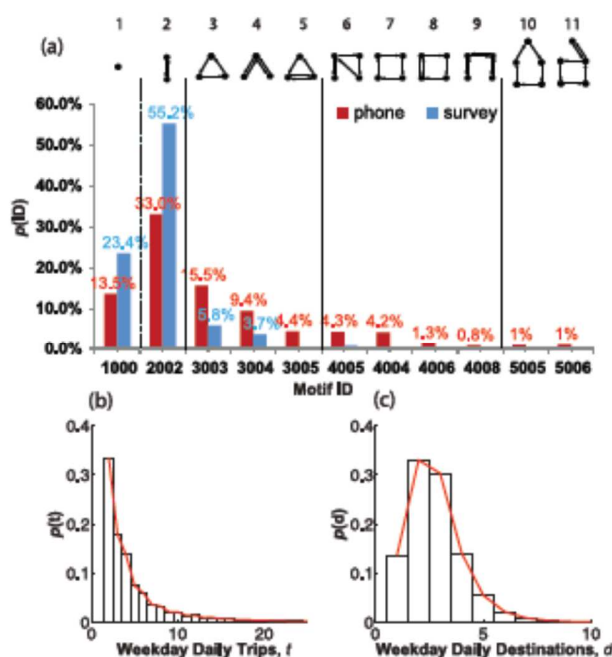
• $\theta_{ik} = \beta_i / K_i$ ، که در آن K_i تعداد کلی روزهاست زمانی که نقوش کاربر i شناسایی می‌شود.

در شکل 8 الف توزیع مکانی عوامل گسترش کاربر در سطح برج در سراسر سنگاپور را نشان می‌دهیم. که توسط گروه‌های مختلف از عوامل گسترش مبتنی بر رتبه‌بندی مقیاس نحراف استاندارد کدگذاری رنگی شده است. می‌توانیم

بینیم که عوامل گسترش کاربر در بخش شمالی سنگاپور بیشتر است، بدان معناست که کاربر تلفن به نسبت جمعیت در این منطقه کمتر از میانگین شهر است. با اینحال زمانی که نگاهی به شکل 4 می‌اندازیم می‌بینیم که برای یک کاربر تلفن همراه متوسط در این منطقه شمالی، وی تمایل دارد تا دارای روزهای نسبتاً بیشتری از مشاهدات باشد. در منطقه مرکزی، عوامل گسترش کاربر کمتر است، این بدان معناست که کاربر تلفن به نسبت جمعیت در این منطقه بالاتر از میانگین شهر است. شکل‌های 8 و 8ج نشان‌دهنده‌ی توزیع فراوانی عوامل گسترش کاربر (برای 1.55 میلیون کاربر تلفن همراه) و عوامل گسترش روز کاربر (برای 6.28 میلیون مشاهده روز کاربر) است.

4.5.4. برآورد جمعیت با وضوح بالا

با عامل گسترش برآورد شده برای هر کاربر تلفن در مطالعه، پس می‌توانیم جمعیت را با وضوح بالا برآورد کنیم که در آمارهای سرشماری موجود نیست اما برای برنامه‌ریزی شهری و حمل‌ونقل هفید است. شکل 9 الف تراکم جمعیت مسکونی (برای افراد بالای 10 سال) در سنگاپور در سطح MTZ (با بیش از 100 منطقه) را با استفاده از داده‌های CDR ارائه می‌کند. برای مقایسه برآورد جمعیت از کاربرانی که تنها دارای محل خانه شناسایی شده (در برآورد 1 ذکر شده است) هستند، با برآوردهایی از کاربران فیلتر شده که الگوهای تحرک آنها را می‌توان شناسایی نمود (ذکر شده در برآورد 2)، دو برآورد جمعیت در



شکل 10. (الف) نقوش روزانه بدست آمده از داده‌های بررسی و CDR، و (ب) توزیع فراوانی سفرهای روزانه، و (ج)

مقاصد فعالیت از داده‌های CDR.

را در سطح برج و در سطح MTZ به ترتیب در شکل‌های 9 ب و 9 ج ترسیم می‌کنیم. می‌توانیم ببینیم که با قلمداد کردن و القای مناسب گسترش، تنها با استفاده از مشاهدات مکرر کاربر (هر چند اندازه نمونه کوچک‌تر باشد)، این می‌تواند برآوردهای قیاس‌پذیری را با برآوردهای بدست آمده با استفاده از تمامی مثال‌های کاربر ارائه کند که محل خانه آنها را می‌توان بدون فیلترینگ (بر اساس فراوانی استفاده از تلفن همراه آنها) شناسایی کرد. مزیت استفاده از کاربران تلفن فیلتر شده (بر اساس فراوانی ارتباطاتی تلفن آنها) در این است که می‌توان الگوهای تحرک غنی‌تر را از سوابق آنها کاوش کرد، و این مشاهدات را بعنوان بررسی سفر در مقیاسی بزرگتر تبقی نمود، تا برنامه‌ریزی درستی را درباره شهر طراحی نمود و همچنین تصمیمات خردمندانه‌ای را درباره سیاست‌های عمومی برای توسعه شهری و حمل‌ونقل گرفت.

4.5.3 جمع‌آوری سفرهای روزانه و الگوهای تحرک

پس از به کار بردن عوامل گسترش برای مشاهدات روز کاربر، سفرهای روزانه و الگوهای تحرک برای ساکنین سنگاپور (بالای 10 سال) را در سطح کلانشهری جمع‌آوری می‌کنیم، و الگوهای تحرک انسانی برای سنگاپور را استنتاج می‌کنیم. شکل 10 توزیع فراوانی استنتاج شده از (الف) الگوهای تحرک روزانه، (ب) تعداد مسافرت‌های روزانه، و (ج) تعداد مقاصد مخصوص روزانه را که کاربران تلفن همراه در طول روز بازدید می‌کنند، ارائه می‌کند. همچنین شکل 10 الف نتایج بدست آمده از داده‌های CDR را با برآوردهای حاصله از داده‌های بررسی سفر HITS مقایسه می‌کند. با توجه به داده‌های تلفن همراه، در می‌یابیم که در یک روز میانگین، 13.5 درصد از ساکنی سنگاپور در خانه می‌مانند، 30 درصد از آنها سه مکان هستند، 14 درصد دارای چهار مکان می‌باشند، 5.5 درصد 5 مکان اقامت یا ماندن دارند، و 21 درصد دارای شش مکان هستند، و کمتر از 2 درصد نیز دارای بیش از 6 مکان هستند. این الگوهای 90 درصد از سفر پاسبه دهندگان بررسی را تحت پوشش قرار می‌دهد.

زمانی که برآوردهای نقوش با داده‌های بررسی مقایسه می‌شود، در شکل 10 الف استدلال ما بر این است که موتیف با دو گره گزارش شده در بررسی غالب است (55 درصد)، و بیشتر از موارد مشاهده شده از داده‌های CDR است. به این نتیجه می‌رسیم که این دلایل برای توضیح بسنده می‌کند. (1) بررسی یا نظرسنجی HITS صرفاً درباره سفرهای موتوری (بدون پای پیاده) می‌باشد، و سفرهای غیر موتوری را شامل نیست. (2) مردم تمایل دارند تا در مقایسه با مقاصد اولیه، کمتر مقاصد ثانویه خودشان را گزارش دهند. اگر به داده‌های بررسی سفر در سایر مناطق کلانشهر همچون شیکاگو، بوستون، و پاریس رجوع کنیم، می‌بینیم که الگوهای سفر دو گره همیشه کمتر از 40 درصد هستند [24]، [35].

یک مطالعه مجزا [45] که بررسی مبتنی بر گوشی‌های هوشمند را برای ردیابی و ثبت سفرهای کاربران منحصر بفرد و فعالیت‌ها در سنگاپور را استفاده می‌کند توسط پروژه تحقیقاتی تحرک آتی (FMS) خبر از مشاهدات مشابه می‌دهد. حتی اگر پروژه FMS تنها دارای 387 نفر باشد که به هر دو سوال یا نظرسنجی مبتنی بر مقاله و سفر اعتبارسنجی خودشان در بررسی مبتنی بر اپ گوشی هوشمند پاسخ داده باشند، با اینحال این مشهود است که افراد بطور قابل توجهی سفر خودشان را در بررسی مبتنی بر مقاله گزارش نمی‌کنند. 76.3 درصد از افراد در بررسی مبتنی بر مقاله دو سفر خودشان را گزارش کردند، در حالی که تنها 20 درصد از افراد شناسایی و تأیید شده بودند که دارای دو مسافرت در بررسی مبتنی بر گوشی هوشمند هستند. بنابراین، الگوهای تحرک حاصله از داده‌های CDR تلفن در اینجا در مقایسه با داده‌های بررسی مسافرت معقول‌تر هستند. این تفاوت تأثیر قابل توجهی بر روی پیامدهای سیاست‌گذاری برای برنامه‌ریزی شهری و حمل و نقل دارد.

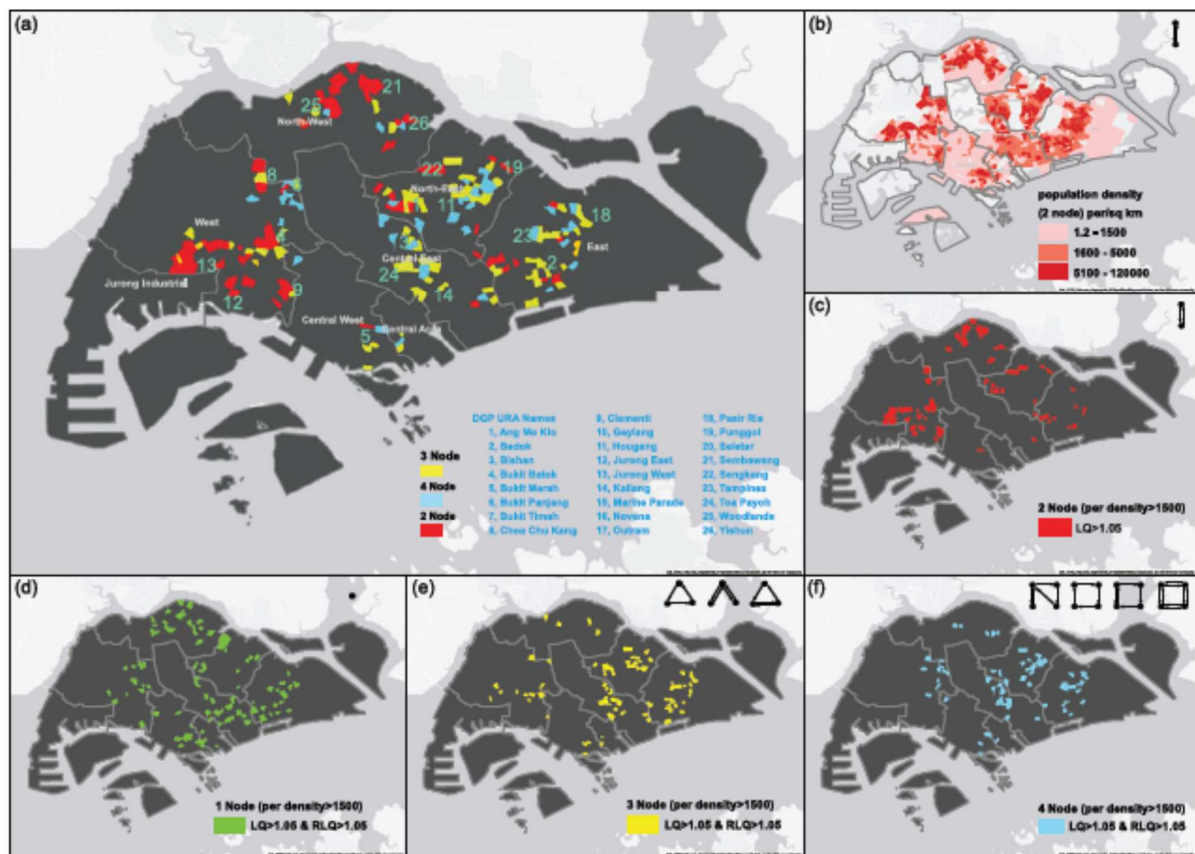
5. الگوهای مکانی تحرک انسانی مبتنی بر فعالیت

درک الگوهای مکانی تحرک انسانی جمع‌آوری شده در محل اقامت آنها بسیار مهم است بدین منظور که به برنامه‌ریزان کمک کند تا دریابند که چگونه محله و محیط ساخته شده آنها ممکن است سفر آنها را تحت تأثیر قرار دهد. مجموعه وسیعی از منابع در برنامه‌ریزی [46] تلاش کرده است تا این تئوری را از نقطه نظر تجربی با استفاده از داده‌های قدیمی بررسی در مورد مسافرت کاوش کند. با روش ارائه شده در این مطالعه، می‌توانیم شواهد و مدارک جدیدی را درباره

رفار مسافرتی افراد در مقیاس بزرگ به دست آوریم، و برنامه‌ریزان و سیاست‌گذاران را تسهیل دهیم تا مناطقی با پتانسیل حجیم همچون ارائه گزینه‌های حمل‌ونقل برای کاهش سفرهای حرکتی، بهبود تسهیلات اجتماعی برای جمعیت مورد هدف، را شناسایی کنند. در این بخش، شاخص‌های تحلیلی را به منظور اندازه‌گیری توزیع مکانی الگوهای تحرک از منظر محل اقامت افراد را ارائه می‌کنیم.

5.1. گردش مبتنی بر خانه دو گرهِه

شکل 11 ب تراکم جمعیتی ساکنان با نقوش دو گرهِه را نشان می‌دهد. این ساکنین یک جای اضافی در کنار «خانه» دارند که در زندگی روزمره آنها حائز اهمیت است. این محل کاری برای کارگران، مدرسه برای دانش‌آموزان، یا محلی اجتماعی / تفریحی برای افرادی است که نه دانش‌آموز هستند و نه کارگر. از آنجا که آنها اکثریت جمعیت (حدوداً 33 درصد، همانطور که در شکل 10 الف نشان داده شده است) را تشکیل می‌دهند، توزیع مکانی گروه با نقوش دو گرهِه بسیار مشابه با توزیع مکانی کل جمعیت (نشان داده شده در شکل 9 الف) است. مفهوم خارج قسمت موقعیت را استفاده می‌کنیم، تا توزیع منطقه ساکنین را توسط نوع تحرک به توزیع از یک ناحیه مرجع (یعنی، منطقه کلانشهر سنگاپور) مقایسه کنیم. برای مثال، برای اشاره دقیق‌تر به موضوع، اندازه‌گیری و مقیاسه می‌کنیم که چگونه ساکنین با موتیف روزانه m به یک محله مخصوص (در سطح MTZ) در مقیاسه با جمعیت در منطقه کلانشهر متصل می‌شوند، در زیر از خارج قسمت محلی (LQ_m) استفاده می‌کنیم.



شکل 11. مناطق خیلی متمرکز با ساکنین (ج) دو گرهمه، (د) تک گرهمه (در خانه مانده)، (و) سه گرهمه، (ی) چهار گرهمه، (ن) پوشش شبکه‌های تحرک دو گرهمه، سه گرهمه، و چهار گرهمه؛ و (ب) تراکم جمعیت ساکنین با نقوش دو گرهمه. (توجه: جمع‌آوری بر اساس محل اقامت کاربران تلفن همراه است).

• که در آن $LQ_m = (e_m/e)/(E_m/E)$ جمعیت کسونی با نوع موتیف روزانه m (شکل 9 الف) در یک منطقه (بعنوان مثال MTZ) است؛ e جمعیت مسکونی در منطقه است؛ E_m جمعیت مسکونی با موتیف روزانه نوع m در منطقه کلانشهر (بعنوان مثال سنگاپور) است؛ E کل جمعیت ساکن در منطقه کلانشهر است.

با توجه به شکل‌های 11 الف و 11 ج، می‌توانیم ببینیم که چندین محله حومه هویت منحصربفردی با بیشتر از سهم متوسط کلانشهری ساکنین دارند که شبکه‌های تحرک روزانه آن دو گرهمه است. این نواحی شامل مناطقی در نواحی غربی، همچون ژورانگ شرقی و غربی، کلمنتی، بوکیت باتوک، چوآچوکانگ؛ مناطقی در منطقه شمال غربی، از جمله وودلندز، سمباوانگ، و بیشون؛ مناطقی در نواحی شمال شرقی همچون آنموکیو، سنگانگ و پانگال؛ و مناطقی در بدوک

منطقه شرقی است (برای منابع نامی منطقه شکل 11 الف را ببینید). افرادی که همدیگر را در کنار خانه خودشان (برای کار/مدرسه بعنوان اکثریت) ملاقات می‌کنند اکثراً مربوط به این مناطق هستند، که شهرهای جدید و همچنین دور از مرکز است. این امر بطور بالقوه برای ارائه خدمات حمل‌ونقل عمومی با ظرفیت و مقدار بالا در طول ساعت پیک حائز اهمیت است تا تراکم را برای ساکنین در این محله‌ها کاهش دهد. همچنین، این موضوع مهم است که خدمات اجتماعی و اقتصادی محلی‌سازی شده‌تری در این محله‌ها طرح‌ریزی شود تا به کاهش سفرهای طولانی مدت برای نیاز به فعالیت‌های اضافی کمک شود.

5.2. نقش تک گره، سه گره و چهار گره

همانطور که الگوی سیاحت مبتنی بر خانه دو گره (شکل 10 را ببینید) کثرت الگوهای تحرک در منطقه کلانشهری است، پس آن را بعنوان یک مرجع استفاده می‌کنیم، و سایر انواع الگوهای تحرک را برای توزیع مکانی الگوی دو گره استفاده می‌کنیم، و آن را خارج قسمت موقعیت نسبی می‌نامیم، برای نقوش نوع d ، $RLQ_d = LQ_d/LQ_2$. شکل‌های 11d، 11ه، و 11و بر MTZها تاکید دارند که غلظت یا مقدار بالائی از ساکنان مقیم در خانه، و ساکنینی با سه گره، و چهار گره در نقوش فعالیت آنها را دارد. شکل‌های 11د و 11ه توزیع مکانی نقوش مختلف با سه گره، و چهار گره در یک نقشه را به تصویر می‌کشد. شکل 11 تنها مناطقی با جمعیت (برای نوع مربوطه از نقوش) بیشتر از 500 نفر و تراکم جمعیت 1500 نفری در هر کیلومتر مربع را نشان می‌دهد. همچنین تنها مناطقی را نشان می‌دهیم که خارج قسمت موقعیت آنها، و خارج قسمت موقعیت نسبی (با دو گره) بیش از 1.05 (مجاز برای خطای 5 درصدی) باشد، تا نواحی بسیار متمرکز با انواع نقوش روزانه مرتبط برای ساکنین مقیم در نواحی را نشان دهیم. شکل 11 الف مناطق بسیار متمرکز از ساکنین با نقوش دو گره، سه گره، و چهار گره در رنگهای مختلف را تحت پوشش قرار می‌دهد، و الگوی کلی تحرک ساکنین سنگاپور را نشان می‌دهد. می‌توانیم ببینیم که ساکنینی که هر روز در خانه می‌مانند



شکل 12. محله‌هایی با غلظت بالایی از ساکنین با نقوش سه گرهِه و چهار گرهِه (با ناحیه‌ای که به اشامی در شکل

11 الف اشاره دارد)

در مناطق در نواحی مرکزی و در امتداد سواحل شرقی، و در نواحی شمال غربی، و برخی از مناطق در نواحی غربی بسیار متمرکز هستند. در نواحی مرکزی، این شامل محله‌های کاملی است که در آنها ساکنین قدیمی از احتمال بالایی برای اقامت برخوردار هستند.

شکل 12 بر روی محله‌هایی زوم می‌کند که حاوی ساکنین متمرکزتر با نقوش سه گرهِه و چهار گرهِه نشان داده شده در شکل 11 الف است. می‌بینیم که این نواحی شامل حومه‌هایی است که توسط ریل سبک (خطوط خاکستری رنگ در شکل) به خطوط اصلی ترانزیت، همچون باکیت پانجانگ، و باکیت باکوک در منطقه غرب (شکل 12)، و سنگانک در منطقه شمال- شرق (شکل 12 ب) متصل می‌شود، که در آن قابلیت دسترسی نسبتاً پائین است. همچنین شامل مناطقی است که در حال حاضر توسط MRT به کار گرفته نمی‌شوند، مانند لنتور، میفلور، برایت هیل و آپر توماسون، در آنگ مو کیو و بیشان (شکل 12 ج)، و محله‌هایی در بخش شمالی بدوک و تمپینس (شکل 12 د). محله‌هایی که

غلظت بیشتری از ساکنین نقوش سه گرهه و چهار گرهه دارند معمولاً بیشتر مسافرت می‌کنند. با ارائه خدمات حمل‌ونقل عمومی بسیار راحت، سطح بهبود یافته‌ای از خدمات (بعنوان مثال برای خط آهن سبک)، و جایگزین نمودن حالات حمل‌ونقل می‌توان کمک به کاهش کل مسافرت کرد. شایان ذکر است که در برخی از مناطق، LTA سنگاپور طرح‌هایی را برای بهبود شبکه ترانزیت ارائه می‌کند (و برخی از آنها تحت ساخت هستند همانطور که در خط نقطه‌چین خاکستری در شکل‌های 12 ج و 12 د نشان داده شده است).

6. نتیجه‌گیری

برنامه‌ریزی شهری و حمل‌ونقل پایدار تا حد زیادی به درک الگوهای تحرک در مناطق کلانشهری بستگی دارد. در این مقاله، خط لوله ادغام یافته‌ای را ارائه می‌کنیم که داده‌های CDR تلفن همراه انبوه و منفعل خام را تجزیه، فیلتر، و گسترش می‌دهد تا الگوهای تحرک انسان برای میلیون‌ها انسان ناشناس ساکن در منطقه کلانشهری را استخراج کند، و دانش دانه‌ای را به زبان‌های قابل تفسیر - برنامه‌ریزی برگردان کند. داده‌های بررسی مسافرت سنتی، هرچند غنی در جزئیات، می‌تواند همراه کننده باشد. این ممکن است الگوهای تحرک غیر دقیق را با توجه به خودگزارشی ناقص تولید کند، و منتهی به مدل‌های تقاضای مسافرت بخصوص برای اهداف سفر بدون کار شود، و منتج به تخصیص منابع ناکارآمد و طرح‌های ناآگاه شود. داده‌های بزرگ، اگر به درستی تلقی شوند، درک بیشتری را ورای بررسی‌های سفر ارائه می‌کنند، هم چنان که مشاهدات چند روزه را پشتیبانی می‌کنند، الگوهای تحرک قدرتمندتری را نشان می‌دهند، و نواحی جغرافیائی وسیع‌تری را تحت پوشش قرار می‌دهند. روش ارائه شده در این مقاله متفاوت از مطالعات موجود است که داده‌های CDR را استفاده می‌کند تا سفرهای انسان در رویکرد مبتنی بر گردش را بررسی کند [26]، [28]. بجای آن، این رویکرد مبتنی بر فعالیت است و متمرکز بر الگوهای گردش و رفتار سفر زنجیره‌ای در شبکه‌های تحرک روزانه است.

همانطور که منابع نشان می‌دهد، با توجه به ساختار اقتصاد شهری در حال تغییر، مشارکت نیروی کار متنوع، و زمانبندی کاری منعطف، رفتار زنجیره‌ای مسافرت بسیار پیچیده می‌شود. سیاست‌ها برای کاهش ازدحام ساعت پیک یا کاهش VMT که افزایش نیاز به مسافرت‌های بدون فعالیت زنجیره‌ای در تردد سفرها یا تمرکز بر بهبود یافتگی‌های

سرعت مسافرت یا هزینه سفر را نادیده می‌گیرد، پس انگیزه‌های می‌تواند کمتر از حد انتظار باشد [47]. علاوه بر این، روش ارائه شده در اینجا بیشتر توسعه می‌یابد تا جزئیات بیشتری را در خوشه‌ها با یک یا دو گره در برابر سه یا چهار نقوش گره بررسی کند - از جمله درک جمعیت شناسی، انواع خانه‌داری خوشه‌ها، که برای تحقیقات سطح بعدی خواسته می‌شود که نقوش توسط طول سفر ترتیب‌بندی می‌شود. با انواع ابزار ارائه شده در اینجا، برنامه‌ریزان مدارک جدیدی را درباره رفتار مسافرتی مردم در مقیاسی بزرگ با استفاده از داده‌های همه جا حاضر از فن‌آوری‌های ارتباطاتی فراهم می‌آورند. با درک الگوهای مختلف تحرک انسانی برای محله‌های مختلف در مقیاس کلانشهر، برنامه‌ریزان سیاست‌هایی را طراحی می‌کنند که مجاوران را در برابر پتانسیل وسیع برای بهبود هدف قرار می‌دهد. شواهد رفتاری در مقیاس بزرگ می‌تواند برای ترتیب‌بندی خوشه‌های خرید در کنار محل‌های خاص برای توسعه حمل‌ونقل محور مفید باشد، و جایگزین‌های حمل‌ونقل را اضافه کند و سطح خدمات را همراه با مسیرهای خاصی بهبود بخشد. سایر کاربردها از جمله بهبود امکانات اجتماعی برای جمعیت مورد هدف، یا ادغام برنامه‌ریزی حمل‌ونقل استفاده از زمین کل سفر و مزایای محیطی و اجتماعی منفی آن را کاهش می‌دهد.



این مقاله، از سری مقالات ترجمه شده رایگان سایت ترجمه فا میباشد که با فرمت PDF در اختیار شما عزیزان قرار گرفته است. در صورت تمایل میتوانید با کلیک بر روی دکمه های زیر از سایر مقالات نیز استفاده نمایید:

لیست مقالات ترجمه شده ✓

لیست مقالات ترجمه شده رایگان ✓

لیست جدیدترین مقالات انگلیسی ISI ✓

سایت ترجمه فا ؛ مرجع جدیدترین مقالات ترجمه شده از نشریات معتبر خارجی