



ارائه شده توسط:

سایت ترجمه فا

مرجع جدیدترین مقالات ترجمه شده

از نشریات معتبر

نویدهای داده های بزرگ و داده های کوچک برای تحلیل رفتار سفر (معادل

تحرک انسانی)

چکیده

دهه ی گذشته شاهد پیشرفت بسیار فعالی در دو حوزه ی وسیع اما مجزا بوده است، که هم شامل شناخت و هم مدلسازی شیوه ای میشود که افراد در زمان و مکان حرکت میکنند (زین پس "تحلیل رفتار سفر" یا "تحلیل تحرک انسان" نامیده میشود). یک رشته که شامل محققان حمل و نقلی میشود که در این رشته برای چندین دهه کار کرده اند و دیگری تازه واردانی را در بر میگیرد که از طیف وسیعی از گرایشها میباشند، اما در اصل دانشمندان کامپیوتر و فیزیکدانان هستند. محققان در این دو رشته با مجموعه داده های مختلف کار میکنند، متدولوژی های مختلفی را بکار میبرند و به سوالات مختلف اما دارای وجه اشتراک پاسخ میدهند. دیدگاه ما این است، لازم است همکوشی پنهان زیاد بین این دو رشته از اختفا بیرون آید. به این ترتیب هدف این مقاله معرفی مجموعه داده ها، مفاهیم، دانش و روشهای بکار رفته در این دو رشته است و مهم تر از همه اینکه ایده های بین گرایشی را برای گفتگوها و همکاری های بین این دو ایجاد میکنند. امید ما بر این است، این مقاله بسیاری از مطالعات متقاطع آینده را برانگیزد که شامل محققانی از هر دو رشته میشود.

کلید واژه ها: داده های بزرگ، داده های کوچک، تحرک انسان، رفتار سفر، برنامه ریزی حمل و نقل

محتویات

1. مقدمه

2. داده های بزرگ

2.1 ماهیت داده های تلفن همراه بی نام ایجاد شده به طور منفعل

2.1.1 انواع داده

2.1.2 تعداد سوابق

2.1.3 تفکیک زمانی

2.2 متدولوژی های فعلی در تحقیقات داده های بزرگ

2.2.1 پیش پردازش

2.2.2 از ردیابی ها تا مکانهای فعالیت

2.2.3 استنباط مکانهای فعالیت (اهداف سفر)

2.2.4 استنباط انتخابهای حالت و مسیر

2.3 مسائل حل نشده (برای کاربردهای برنامه ریزی حمل و نقل)

2.3.1 اعتباریابی

2.3.2 معرف بودن

3. عوامل رفتاری

4. توسعه مدل

5. الگوهای تحرک انسانی

6. بحثها

تقدیرنامه ها

منابع

1. مقدمه

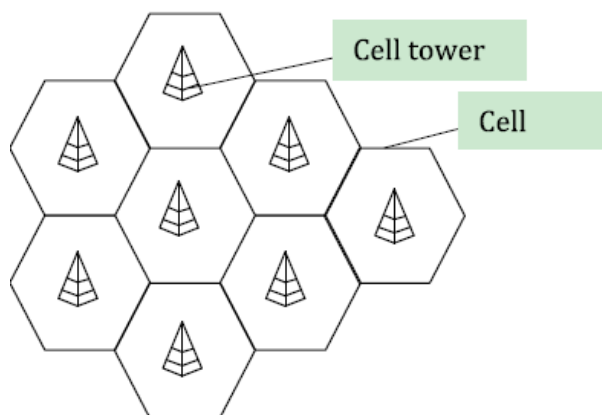
تحقیقات در زمینه ی حرکت انسان در زمان و فضا، حداقل به مدت پنج دهه وجود داشته است. این تحقیقات با نیاز به پیش بینی تقاضای سفر آینده به منظور هدایت بهتر سرمایه گذاری اغلب پروژه های حمل و نقل در مقیاس وسیع برانگیخته شده اند، محققان حمل و نقل از دیر باز در صدد توسعه ی مدلهایی برای پیش بینی شیوه ای هستند که مردم در زمان و فضا سفر میکنند و در جستجوی شناخت عواملی هستند که بر انتخاب های مربوط به سفر اثر میگذارند. اخیراً، چالشهای عظیمی هم چون گرم شدن جهانی و آلودگی هوا ممکن است همگی به علت اتکاء بیش از حد به اتومبیلها باشد، به علاوه محققان و کارشناسان حوزه ی حمل و نقل را برای توسعه ی راهبردهای موثر برای حرکت به سوی حالات نگهداشت پذیرتر حمل و نقل بر می انگیزاند (مثلاً

ترانزیت عمومی و پیاده روی و دوچرخه سواری). چندین دهه است که محققان حمل و نقل به طور وسیعی از داده های تقاضای فعال استفاده نموده اند من جمله به عنوان مثال زمینه یابی های سفری که در آن از افراد خواسته میشود فعالیتهای و سفرهای خود را از طریق مصاحبه های کتبی، شبکه ای یا تلفنی، شخصا گزارش دهند؛ زمینه یابی های سفر به واقعه نگارهای جی پی اس (GPS) متصل شده اند که در طول آن از افراد خواسته میشود هم پرسشنامه ها را تکمیل کنند و هم واقعه نگارهای جی پی اس را حمل نمایند؛ و زمینه یابی های مبتنی بر جی پی اس خالص که در طول آنها از افراد فقط خواسته میشود واقعه نگارهای جی پی اس را حمل کنند. در نوع آخر، اطلاعات در مورد فعالیتهای و سفرهای افراد باز هم لازم است از ردیابی های جی پی اس جمع آوری شده، استنباط شود. کل این زمینه یابی ها دارای مشخصات مشترکی هستند و آن بدین شرح است: تقاضای فعال- افراد و اطلاعات در مورد سفرهایشان به طور فعالی بکار گرفته میشوند. احتمالا به خاطر این ویژگی، این زمینه یابی ها با اندازه ی نمونه ی نسبتا کوچکی محدود میشوند. در این مقاله، ما به داده های تقاضای فعال اشاره میکنیم.

به موازات استفاده ی مداوم از داده های کوچک در تحقیقات حمل و نقل، پیشرفت و رواج سریع فناوریهای موبایل، جمع آوری میزان وسیعی از داده های منفعل (داده های بزرگ) را میسر ساخته است که منجر به موج خروشان مطالعات در مورد حرکت انسان شده است. داده های منفعل به داده های بکار رفته برای تحقیقات اشاره دارد. نمونه ها شامل مشاهدات تلفن همراه ایجاد شده از سوی اپراتورهای همراه برای اهداف عملیاتی، داده های رسانه های اجتماعی ایجاد شده به طور داوطلبانه توسط فعالیتهای آنلاین کاربران و داده های کارتهای هوشمند جمع آوری شده در بسیاری از سیستم های ترانزیت در سطح جهان میباشد. چنین داده هایی که به صورت منفعل جمع آوری شده اند، با داده های تقاضای فعال (داده های کوچک) بسیار متفاوت هستند که برای اکثر محققان حوزه ی حمل و نقل، آشنا هستند و به این ترتیب به روشها و تکنیکهای متفاوتی برای پردازش و مدلسازی نیاز دارند. هدف اول این مقاله، معرفی داده های بزرگ جمع آوری شده به طور منفعل برای محققان حمل و نقل است، که یک بررسی مدرن از روشهای بکار رفته را ارائه داده و نواحی دارای شکاف را شناسایی میکند که به طور خاصی برای برنامه ریزان حمل و نقل، حائز اهمیت میباشد.

از آن مهم تر اینکه، این مقاله در صدد شناسایی مفاهیم میان گرایشی و فرصتهایی برای هم محققانی است که دارای داده های کوچک بکار رفته به طور متداول هستند و هم محققانی که دارای داده های بزرگ میباشند. بحث ما، در سه بخش فرعی مهم تحقیقات رفتار سفر خواهد بود: 1) عوامل رفتاری که در آن علاقه عواملی را شناسایی میکند که رفتارهای سفر را شرح میدهد و مکانیسم های سببی متضمن را آشکار مینماید؛ 2) مدلسازی رفتار سفر که در آن مدلها برای پیش بینی رفتارهای حرکت انسان ایجاد میشوند؛ و 3) الگوهای تحرک انسانی که در آن الگوشناسی، یک هدف مهم محسوب میگردد. دیدگاه ما این است که تحرک انسانی و به طور همزمان، دانش انباشته شده در تحقیقات حمل و نقل در چندین دهه ی گذشته میتواند مطالعات داده های بزرگ را برای پاسخ دهی به سوالاتی هدایت نماید که برای جامعه مهم است به طور خاص آنهایی که به تصمیمات و خط مشی های سرمایه گذاری حمل و نقل در محیطهای شهری مربوط هستند.

بقیه این مقاله به صورت زیر برنامه ریزی میشود. در بخش 2، مقدمه ای از داده های بزرگ هم چنین نقد و بررسی متدولوژی های کنونی ارائه میشود. تاکید ما بر مجموعه داده های تلفن همراه ایجاد شده به طور منفعل است. در تحلیل الگوهای تحرک انسان، داده های تلفن همراه ایجاد شده به طور انفعالی به صورت مکررترین منبع داده ی مورد استفاده (و شاید قابل اعتمادترین) پدید آمده است. سایر منابع داده ها نمیتوانند طیف کاملی از یک الگوی تحرک انفرادی را در روزهای متعدد نشان دهند، من جمله استفاده از حالات متعدد حمل و نقل. نمونه ها عبارتند از استفاده از داده های تاکسی که تقریبا برای مطالعه ی الگوهای رانندگان، در جستجو برای مسافری مناسب است، استفاده از داده های کارت هوشمند ترانزیت که فقط استفاده از حالات عبور را نشان میدهد، یا استفاده از داده های رسانه های اجتماعی که تفکیک مکانی و زمانی شان از داده های تلفن همراه بسیار پایین تر است و به سوی موقعیت های مکانی خاصی متمایل است. جماعت هدف بخش 2، محققان حمل و نقلی است که با داده های زمینه یابی کوچک آشنایی دارند اما با داده های بزرگ آشنایی ندارند که اخیرا مورد استفاده قرار گرفته شده اند. در بخش 3-5 مفاهیم میان گرایشی و ایده ها را در سه بخش فرعی پیشتر ذکر شده مورد بحث قرار میدهم. نتیجه گیری بحثها در بخش 5 ارائه میشود.



شکل 1. یک شبکه سلولی نمونه.

جدول 1: سوابق نمونه در داده های CDR.

X	Y	ID	زمان	مدت (ثانیه)
195925	32464	J000001	82141	81
195925	32464	J000001	82456	75
195018	31555	J000002	82100	140

الف. مختصات XY از دستگاه مختصات جغرافیایی منتقل میشوند. یک تبدیل میتواند برای تبدیل آنها به مختصات طول و عرض جغرافیایی صورت گیرد.

جدول 2: نمونه ای از داده های مشاهدات.

ID	Time ^a	Location ^b
3X35E90	13 19242582	34.044162 -112.454400
3X35E90	13 19242583	34.044059 -112.455550
3X35E90	13 19301785	34.044392 -112.453519

الف. زمان، تاریخزن یونیکس است که به صورت تعداد ثانیه هایی تعریف میشود که از ساعت جهانی هماهنگ 00:00:00، سه شنبه، 1 ژانویه 1970 طی شده است.

ب. موقعیت مکانی، مختصات طول و عرض جغرافیایی تلفنهای همراه است.

2. داده های بزرگ

2.1 ماهیت داده های تلفن همراه بی نام ایجاد شده به طور منفعل

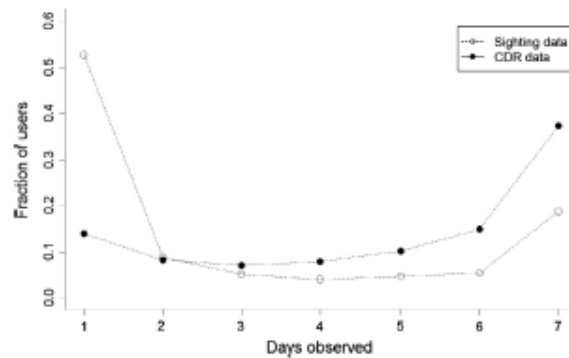
2.1.1 انواع داده ها

اطلاعات موقعیت مکانی در یک مجموعه داده ی تلفن همراه منفعل به صورت نتیجه ی یک ارتباط تلفنی با شبکه ی سلولی حفظ شده و بکار انداخته شده توسط اپراتورهای شبکه ی سلولی ایجاد میشود (مثلا ارائه دهندگان تلفن)؛ این فرایند، موقعیت یابی نامیده میشود. یک شبکه ی سلولی، شبکه ای است که تلفنهای همراه را قادر میسازد با همدیگر ارتباط برقرار نمایند؛ از ایستگاه های متعدد تشکیل میشود، هر یک به یک ناحیه خدمت رسانی میکنند که یک سلول نامیده میشود (شکل 1). هر سلول، دارای یک ID سلولی منحصر به فرد است.

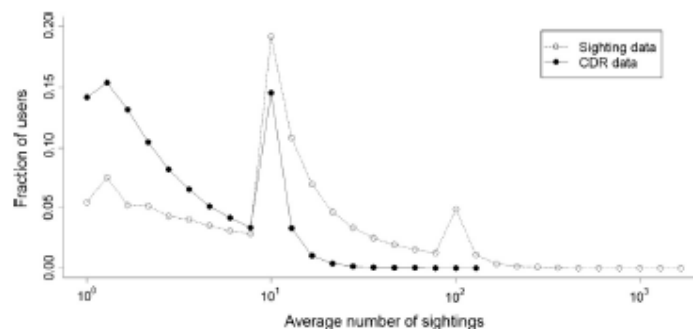
موقعیت یابی تلفن همراه زمانی لازم میشود که یک کاربر با شبکه ارتباط برقرار نماید. هنگامیکه کاربری یک رویداد اتصال شبکه را شروع میکند (مثلا تماس صوتی)، اپراتور شبکه ی سلولی لازم است موقعیت مکانی او را برای تعیین برج سلولی بکار رفته برای انتقال این رویداد، بشناسد. لذا، داده های موقعیت یابی (حاوی اطلاعاتی پیرامون موقعیتهای مکانی کاربران) زمانی ایجاد میشود که یک رویداد اتفاق می افتد. چنین داده هایی به طور خودکار و انفعالی برای اهداف خود اپراتورهای شبکه ی سلولی ایجاد میشوند من جمله جمع آوری اطلاعات صدور قبض و مدیریت شبکه.

موقعیت یابی میتواند به دو نوع داده ی تلفن همراه منجر شود. داده های CDR (CDR) احتمالاً امروز، در اکثر موارد بکار میروند (مثلا [Gonzalez et al., 2008](#); [Kang et al., 2012a, 2012b](#)). هر ثبتی در یک داده ی CDR، یک تماس تلفنی را نشان میدهد با اطلاعاتی در مورد تماس گیرنده، کسی که با او تماس گرفته میشود، زمان شروع تماس، مدت زمان تماس و مختصات XY برجی که اولین بار تماس را منتقل کرده است هنگامیکه برای اولین بار تماس شروع شد. یک نمونه از داده های CDR در جدول 1 نشان داده شده است که در آنجا ID به کاربری اشاره دارد که یک تماس را شروع یا دریافت میکند، X و Y، مختصات برجی است که تماس

کاربر را منتقل میکند و TIME و DURATION



شکل 2. تعداد روزهایی که در آنها حداقل یک ثبت در یک روز واحد مشاهده شده است.



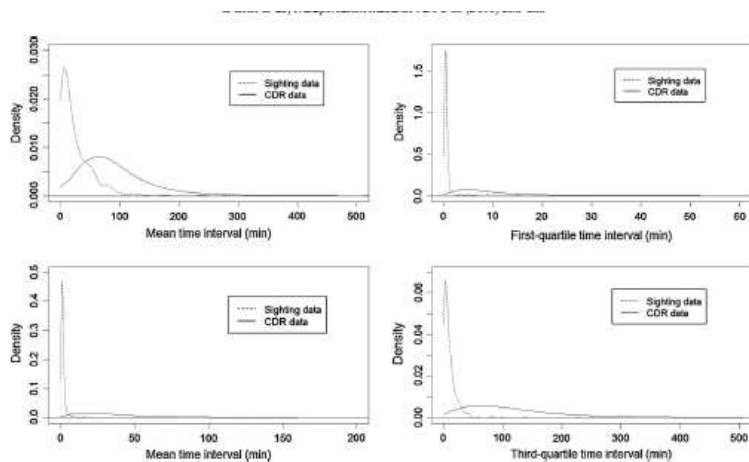
شکل 3. تعداد متوسط ثبتها در روز به ازای هر کاربر.

زمان شروع و طول مدت تماس را ذکر میکنند. به عنوان مثال، اولین تماس نشان میدهد کاربر "J000001" در ساعت 41:21:8 در این موقعیت مکانی (32464،195925) تماسی به مدت 81 ثانیه داشته است. یک مجموعه داده ی تلفن همراه که به مراتب کمتر مورد استفاده قرار میگیرد "داده های مشاهدات" نامیده میشود (مثلا Ma et al., 2012; Calabrese et al., 2013; Chenet et al., 2014). (جدول 2 را ملاحظه نمایید). هر بار که یک تلفن موقعیت یابی میشود، یک مشاهده ایجاد میشود. لذا، هر ردیف در این مجموعه داده، یک مشاهده را نشان میدهد. ممکن است داده های مشاهده به صورت یک نسخه ی تاحدی پردازش شده از داده های CDR تصور شود. این داده ها با داده های CDR از سه نظر فرق دارند: (1) تفکیک زمانی: داده های مشاهدات احتمالاً از یک سطح تفکیک زمانی بالاتر از داده های CDR برخوردارند- در حالیکه یک تماس تلفنی واحد، یک ثبت را در داده های CDR ایجاد خواهد کرد، همان تماس ممکن است مشاهدات متعددی را ایجاد نماید. (2) تفکیک مکانی: داده های مشاهدات احتمالاً دارای سطح بالاتری از تفکیک مکانی نسبت به داده های

CDR هستند. موقعیتهای مکانی گزارش شده در CDR، موقعیتهای مکانی برج تلفن سلولی هستند و به این ترتیب به چگالی شبکه سلولی وابسته اند که از خیلی کم به صورت چند صد م تر در نواحی کلان شهری تا چند کیلومتر در نواحی روستایی فرق دارند (مثلا Calabrese et al., 2013; Chen et al., 2014). از سوی دیگر، این موقعیتهای مکانی گزارش شده در داده های مشاهدات، نتایج مثلث بندی برجهای متعدد است و این موقعیتهای مکانی به طور وسیعی به صورت موقعیت های مکانی دستگاه قلمداد میشوند که بر خلاف موقعیتهای مکانی برج در داده های CDR است؛ 3) تعاملات کاربر: اطلاعات پیرامون تعاملات کاربر میتوانند به طور مستقیم از داده های CDR مشاهده شوند، زیرا هم تماس گیرنده و هم فرد مورد تماس برای یک تماس تلفنی واحد، ثبت میشوند. این مورد برای داده های مشاهدات وجود ندارد.

2.1.2 تعداد ثبت ها

تعداد روزهایی که در طول آن یک تلفن حداقل یک بار در یک روز واحد مشاهده میشود، به شدت بین کاربران مختلف فرق دارد و این مساله برای هر دو نوع داده وجود دارد. شکل 2 برخی مشاهدات در این زمینه را با استفاده از نمونه های مستقل ارائه میدهد: یک نمونه ی CDR و یک نمونه ی مشاهدات. هر دو توزیع، یک شکل U را نشان میدهد.



شکل 4. بازه های زمانی بین ثبتهای متوالی.

میتوان مشاهده کرد ناهمبستگی بسیار زیادی در تعداد دفعات متوسطی وجود دارد که یک تلفن رویت میشود (داده های مشاهدات) یا یک تماس تلفنی صورت میگیرد (داده های CDR)، همانطور که در شکل 3 نشان داده شده است. برای هر دو مجموعه داده، یک پیک در حدود ده ثبت وجود دارد؛ روی هم رفته (بدون این پیک)،

توزیع یک الگوی نمایی را نشان میدهد، با حداکثر کاربرانی که دارای تعداد ثبت کمتر از 10 مورد هستند و تعداد کمی از کاربران دارای تعداد زیادی ثبت میباشند.

2.1.3 تفکیک زمانی

به طور کلی، انتظار می‌رود داده های CDR از داده های مشاهدات از نظر تفکیک زمانی شان، ناهموارتر باشند. در یک مقاله، داده های CDR، دارای یک زمان بین-رویداد 8.2 ساعت برای 100.000 نفر در یک دوره ی شش ماهه یافت شدند. با استفاده از یک مجموعه داده ی مشاهدات، کالابریست و همکارانش یک زمان بین رویداد 260 دقیقه ای را یافتند، که از زمان بین رویداد در اثر گونزالز و همکارانش بسیار پایین تر بود. به علاوه بازه ی زمانی بین مشاهداتی متوالی را با چارکهای اول، دوم و سوم توصیف نمودند. این مولفین گزارش کردند میانگین حسابی میانه ها به صورت 84 دقیقه است و پی بردند تفکیک زمانی داده هایشان برای کشف تغییرات موقعیت مکانی که کاربر در آنجا برای مدت زمان به کوچکی 1.5 ساعت توقف دارد، به قدر کافی مناسب است. شکل 4 توزیعات بازه های زمانی بین ثبتهای متوالی را نشان میدهد که بر حسب میانگین، میانه، چارکهای اول و سوم برای دو نمونه ی مستقل بیان شده است. الگوهای اوج دهندگی دو مجموعه داده با همدیگر فرق دارند، که برخی الگوهای منحصر به فرد وابسته به دو نوع داده را بازتاب میکند- برای داده های مشاهدات، مشاهدات ایجاد شده میل دارند با هم دسته بندی شوند زیرا یک تماس تلفنی واحد، ایجاد مشاهدات متعدد را مورد هدف قرار خواهد داد؛ برای داده های CDR، این بازه ها به درستی، بازه های بین تماسهای متوالی را منعکس میکنند.

2.2 متدولوژی های فعلی در تحقیقات داده های بزرگ

2.2.1 پیش پردازش

در هر موقعیت مکانی فرضی در یک شبکه ی سلولی، ممکن است چندین برج سلولی وجود داشته باشد که سیگنالهای رادیویی شان به یک دستگاه برسد. اگر این برجهای سلولی متعدد دارای قدرت سیگنال مشابه باشند، اتصال یک دستگاه ممکن است بین چندین برج متعدد جهش داشته باشد حتی هنگامیکه دستگاه ساکن باشد. در این صورت ممکن است به نظر برسد کاربر فقط در عرض چند ثانیه، چندین کیلومتر سفر میکند. این پدیده به صورت نوسان در یک شبکه ی سلولی شناخته میشود. آنچه در داده ها بازتاب شده است، یک دنباله از ثبتهایی است که بین چندین برج جهش داشته اند. تعدادی روش وجود دارد که برای بررسی مساله ی نوسان

ارائه شده اند. روش اتصال مبتنی بر سرعت، از همه بیشتر مورد استفاده قرار میگیرد و به صورت زیر عمل میکند: چنانچه برج الف بین چندین ثبت برج ب ثبت گردد و سرعت راه گزینی بین الف و ب از یک آستانه ی از پیش تعیین شده بزرگتر باشد، آنگاه نوسان کشف میشود. این روش بر مبنای مشاهده ای است که نوسان به تغییر موقعیت مکانی توصیف شده با یک سرعت بالای غیر عادی منجر میشود. چالش این روش، انتخاب یک آستانه ی سرعت است که بین سرعتهای نرمال و غیر نرمال تمایز قائل میشود. متناوبا، یک روش مبتنی بر الگو در برخی مطالعات بکار رفته است. این روش، الگوی منحصر به فرد در بروز رسانی های موقعیت مکانی وابسته به راه گزینی های مکرر- نوسان بین جفتهای موقعیت مکانی را به رسمیت میشناسد. لی و هو (2006)، بروز نوسان را به صورت زمانی تعریف کردند که سه راه گزینی متوالی بین یک جفت موقعیت مکانی، مشاهده شود. زمانیکه نوسان شناسایی شود، کل موقعیتهای مکانی مشمول در این راه گزینی ها، با موقعیت مکانی در جفتی که کاربر بیشتر وقتها به آن وابسته بوده است، جایگزین میشود. یک روش مشابه توسط بایر و همکارانش (2010) بکار رفت. این روش مبتنی بر الگو، دارای خطر اشتباه کردن حرکات واقعی کاربری است که مکررا بین دو موقعیت مکانی برای نوسان سفر میکند. وانگ (2014) یک روش هیبریدی را ارائه کرده است که به صورت زیر عمل میکند: اولاً، زیر دنباله هایی که ظاهراً از نوسان حاصل میشوند، بر مبنای رویکردهای مبتنی بر الگو کشف میشوند؛ آنگاه سرعتهای راه گزینی بین جفتهای موقعیت مکانی برای هر زیر دنباله مشخص میشود؛ در نهایت، زیر دنباله ها فقط به روز میشوند چنانچه سرعت راه گزینی از یک آستانه ی سرعت فراتر باشد همانطور که در رویکردهای مبتنی بر سرعت مشخص شده است.

2.2.2 از ردیابی ها تا مکانهای فعالیت

عدم اطمینان مکانی وابسته به داده های بزرگ (یا داده های CDR یا مشاهدات) بدین معنی است که به خودی خود، ردیابی های موقعیت مکانی در داده ها، موقعیت های مکانی بازدید شده یا مورد عبور را نشان نمیدهند. به بیان دیگر، این ردیابی ها لازم است برای انواع مختلف موقعیتهای مکانی فعالیتی که افراد می بینند، پردازش، تجمیع و استخراج شوند.

به بیان کلی، دو رویکرد اصلی برای تجمیع ردیابی های موقعیت مکانی وجود دارد. اولین رویکرد، برچسب گذاری موقعیتهای مکانی فعالیت بر حسب فراوانی است. این رویکرد فقط برای داده های CDR قابل اجراست که در آنها

تعداد برجهای تلفن سلولی در یک ناحیه، محدود است. برای نمونه، موقعیت مکانی خانه شناسایی میشود چنانچه تلفنی در طول شب (مثلا از 9 شب تا 9 صبح) در یک فراوانی فراتر از یک آستانه ی از پیش تعیین شده موقعیت یابی شود (مثلا 70٪).

روش فراوانی نمیتواند برای داده های مشاهده بکار رود، زیرا یک ثبت موقعیت، حاصل مثلث بندی است و در نتیجه هر ثبتی، منحصر به فرد است. لذا، اغلب خوشه بندی بکار میرود. تعدادی از مطالعات از خوشه بندی مبتنی بر مسافت استفاده میکنند. برای نمونه، یی و همکارانش، ثبتهای موقعیت را در یک مسافت 200 متری خوشه بندی کردند و کالابرس و همکارانش از یک کیلومتر به عنوان آستانه ی مسافت در خوشه بندی استفاده کردند. روشهای خوشه بندی مثل میانگین-K و خوشه بندی سلسله مراتبی به تعداد خوشه ها به عنوان درونداد نیاز دارد. در کاربردهای زیادی، درونداهای لازم توصیف شده در بالا، معمولا غیر قابل دسترسی هستند و موارد انتخاب شده به طور قراردادی، عدم اطمینان به نتایج خوشه بندی را ببار می آورد. استر و همکارانش یک روش خوشه بندی مبتنی بر چگالی را ارائه نموده اند که از چگالی موضعی به جای مسافت برای تعیین خوشه استفاده میکند (1996). این روش به تعداد خوشه ها به عنوان درونداد نیاز دارد و میتواند داده های خارج از محدوده ای را پیدا کند که به هیچ خوشه ای تعلق ندارند. در عین حال، بر روی نقاط مرزی، عدم اطمینانی وجود دارد که ممکن است به خوشه های متعدد تعلق داشته باشد.

الگوریتم انعطاف پذیر دیگری که به تعداد خوشه ها به عنوان درونداد نیاز ندارد، خوشه بندی مبتنی بر مدل است. در این روش، فرض میشود ثبتهای موقعیت از یک مدل آماری ایجاد شوند. مدل ترکیب گائوسی (GMM)، یکی از پر کاربرد ترین مدلها برای آنالیز خوشه بندی محسوب میگردد. یک GMM یک جمع وزنی از توزیعات گائوسی انفرادی متعدد با پارامترهای ناشناخته میباشد و هر توزیع انفرادی به لحاظ ریاضی، یک خوشه را نشان میدهد. برای اجرای این روش، فقط حداکثر تعداد خوشه های M به عنوان یک درونداد لازم است و برای هر تعداد m از 1 تا M ، خوشه های M از طریق الگوریتم بیشینه سازی-انتظار (EM) با معیار اطلاعات بیزی (BIC) برآورد میشود. در نهایت، نتیجه خوشه بندی دارای بالاترین BIC از نتایج M به صورت بهترین، انتخاب میشود. مزیت دیگر این روش این است که نتیجه اش میتواند از طریق احتمالی ارزیابی شود که یک

مشاهده‌ی فرضی به خوشه‌ی تعیین شده‌ی کنونی اش تعلق ندارد. چن و همکارانش، از این روش برای تجمیع ردیابی‌ها و کسب نتایج نوید بخش استفاده نمودند.

2.2.3 استنباط مکانهای فعالیت (اهداف سفر)

اکثر داده‌های بزرگ با یک مساله‌ی "خط سیر داده، غنی، اما اطلاعات فعالیت ضعیف" مواجه هستند. خوشه بندی به موقعیت‌های مکانی فعالیت استنباط شده منجر میشود نه به انواع آنها. برای استنتاج انواع موقعیتهای، محققان از روشهای ساده استفاده نموده اند مثل رویکرد مبتنی بر فراوانی، که موقعیتهای به مراتب بیشتر بازدید شده در طول زمان روز و زمان شب را به ترتیب به صورت موقعیتهای مکانی منزل و محل کار تعیین میکنند یا رویکرد مبتنی بر مدل که انواع موقعیت مکانی را با یک مدل آماری پیش بینی میکنند. به علاوه، اطلاعات کاربری زمین احاطه کننده‌ی یک موقعیت مکانی فعالیت استنباط شده هم چنین میتواند برای تعیین احتمال فعالیت مربوطه بکار رود. ایده‌ی اساسی، پیوند یک موقعیت با POIهای مجاور (نقاط جالب توجه) است، با در نظر گرفتن قواعد تجربی از پیش تعریف شده (مثلا جذابیت POIهای، ساعات خدمات POIها و مدت زمانهای توقف). روشهای مشابه هم به طور وسیعی توسط محققان حمل و نقل با داده‌های جی پی اس بکار رفته اند.

استنتاج ماتریسهای مبدا-مقصد. تخمین ماتریس OD، یکی از مهم ترین گامها در پیش بینی تقاضای سفر، که به طور متداول با مشاهده‌ی مستقیم الگوهای سفر با آماده سازی های سخت افزاری برای نمونه برداری کل یا بخشی از محرکهای وارد شونده یا ترک کننده‌ی خط مرزی مطالعه انجام گرفته است. این فناوریها به یک دسته کاوش فعال تعلق دارد. مشاهده‌ی الگوهای سفر در یک مقیاس بزرگ، فقط با پیدایش داده‌های بزرگ جمع آوری شده به طور منفعل، ممکن شده است، همانند تلفنهای هوشمند و کارتهای هوشمند مثل گذرنامه های ترانزیت و برچسبهای عوارض. پس از استخراج موقعیتهای مکانی فعالیت، برآورد OD شامل محاسبه‌ی زمانهای سفر بین جفت‌های مبداها و مقصدها میشود. این امر شامل نگاشت مبداها و مقصدهای استنباط شده از داده‌های تلفن همراه به شبکه‌ی حمل و نقل میشود (بالا را ملاحظه نمایید). بسته به انواع منابع داده‌ی تلفن موبایل، این نگاشت میتواند با استفاده از رویکردهای مختلف انجام گیرد. چنانچه داده‌های موقعیت مکانی در چند ضلعی ورونوئی برج سلولی ارائه شود (مثل داده‌های CDR)، نزدیک ترین گره شبکه جاده ای یا ایستگاه مترو از چند ضلعی ورونوئی اغلب برای اتصال موقعیتهای مکانی استخراج شده به شبکه انتخاب میشود. اگر داده

های موقعیت مکانی در موقعیتهای طول و عرض جغرافیایی مثلثاتی و پردازش شده باشند (مثلا داده های مشاهدات)، تجمیع به یک ساختار ناحیه از پیش تعیین شده لازم است (مثلا، گریدها، وسعت سرشماری ها یا نواحی آنالیز ترافیک). تجمیع به نواحی ریزتر میتواند به نمایشهای OD نویزی و نامتعادل منجر شود. این مساله زمانی کاهش می یابد که اندازه های ناحیه، بزرگتر میشوند.

برآورد OD هم چنین محاسبه ی عوامل مقیاس گذاری را در بر میگیرد که سفرهای OD استنتاج شده از تلفن همراه را (که هنوز هم مبتنی بر نمونه است) به شماره های میزان- جمعیت تبدیل میکند، گرچه این روش میتواند پیشقدر دار باشد زمانیکه این نسبت بالا باشد (مثلا بیش از 20 برای نواحی خاص دارای مقادیر نفوذ پایین تلفنهای سلولی. این روش شبیه به روش محدود- تکی است که به مبداها یا مقصدها مقیاس گذاری میکند زیرا بر خلاف روش محدود- دوتایی است که در صدد مقیاس گذاری هم تعداد مبداها و هم مقصدها به طور همزمان است. یک روش مقیاس گذاری پیشرفته تر، محاسبه ی عوامل مقیاس گذاری را همراه با مرحله ی تخصیص ترافیک به صورت تکراری یا به صورت یک مساله ی بهینه سازی در بر میگیرد.

2.2.4 استنباط انتخابهای حالت و مسیر

مطالعاتی که انتخابهای حالت و مسیر را از داده های بزرگ استنباط میکنند، بسیار محدودتر هستند، در مقایسه با مواردی که موقعیتهای مکانی فعالیت را کسب میکنند و انواع شان را استنباط میکنند (مطالب فوق را ملاحظه نمایید). استنباط انتخابهای مسیر به طور بارز به دو نوع نقطه ی در دسترس در داده های تلفن همراه متکی است: (1) نقاطی که بیانگر موقعیتهای مکانی فعالیت یا مبداها یا مقصدها هستند؛ و (2) نقاط میانی بین مبداها و مقصدهای شناسایی شده. این نقاط با داده های شبکه ی حمل و نقل پوشانده میشوند (سواره رو یا ترانزیت) تا محتمل ترین انتخاب های مسیر را مشخص کنند. برخی مطالعات از نقاط میانی برای استنباط انتخابهای مسیر استفاده میکنند. این مطالعات نخستین بار یک مجموعه ی انتخاب را ایجاد کردند که از مسیرهای متعدد بین یک مبدا و یک مقصد تشکیل میشود. تعاریف برای معنی کردن چیزی بکار رفته اند که توجیه میکند یک مبدا یا مقصد به شدت با هم فرق دارند من جمله گروه های بلوک سرشماری، چند ضلعی های ورونوئی یا یک ناحیه موقعیت مکانی بزرگ (مثلا LA). این مطالعات هم چنین از متریک های متفاوتی برای اندازه گیری مجاورت به

جاده یا شبکه های ترانزیت بکار رفته اند. در موارد دیگر، این نقاط داده های میانی، بکار نرفتند؛ در عوض یک تکنیک تخصیص ترافیک نموی بکار رفت.

مطالعاتی که انتخابهای حالت را از داده های تلفن همراه استنباط میکنند، حتی از مطالعات برای انتخابهای مسیر هم پراکنده تر هستند. قبل از استنباط انتخابهای حالت، نقاط بدست آمده از داده های تلفن همراه، اولین بار با داده های شبکه پوشانده شدند (مثلا شبکه های ترانزیت یا سواره رو) همانطور که برای انتخابهای مسیر این گونه است. پس از آن، میتوان به عنوان مثال سفرهای هوایی را با ارجاع-جغرافیایی شناسایی کرد که در اطراف موقعیت های مکانی فرودگاه در یک طیف برد معین (مثلا 3 مایل) و با طول مدت سفر معین تمام میشود. یک روش مستقیم تر دیگر، بکارگیری الگوریتم خوشه بندی میانگینهای K کنترل نشده برای تقسیم بندی ثبتها (در داده های تلفن همراه) به سفرهای خودروئی یا ترانزیت از طریق سرعتهای سفرشان میباشد، اگر این نقاط در مبدا و مقصد مشابه واقع شوند.

2.3 مسائل حل نشده (برای کاربردهای برنامه ریزی حمل و نقل)

2.3.1 اعتباریابی

برخلاف زمینه یابی های سفری که در طول آن افراد موقعیت مکانی های بازدید شده خود و حالات حمل و نقل بکار رفته را شخصا گزارش میکنند، مجموعه داده های تولید شده به طور انفعالی دچار کمبود درستی زمین برای معتبر سازی در مقابل آن هستند. احتمالا به این دلیل، فقط تعداد معدودی مطالعه با استفاده از داده های انفعالی این مساله را تاحدی بررسی کرده اند. میزان محدود مطالعات صورت گرفته در این زمینه دچار سفسطه ی اکولوژی هستند – "یک نتیجه گیری در مورد رفتار انفرادی استنباط شده از داده ها در مورد رفتار جمعی" – این مطالعات نتایج استنباط شده را از داده های بزرگ (تجمیع شده بر روی کل کاربران) با نتایج تجمیع حاصل از یک نمونه ی مستقل مقایسه میکنند. در پاره ای از مطالعات، دو نمونه ی مقایسه شده، تناسب ندارند. گرچه یک گام نخست بسیار مهم را به سوی هدایت صحیح نشان میدهند، لازم به ذکر است نتایج استنباط شده در سطح انفرادی میتواند میزان خطای بسیار زیادی داشته باشد گرچه سطح بالایی از صحت و دقت در این سطح انبوه مشاهده میشود.

مطالعات آینده، راه هایی را برای اعتبار یابی نتایج در سطح انفرادی توسعه خواهند داد. به طور بالقوه، ممکن است راه های متعددی وجود داشته باشد. در سال 2014، چن و همکارانش (2014)، روشی را از طریق استفاده از شبیه سازی ها نشان دادند. با استفاده از دو منبع مستقل: یک زمینه یابی سفر خانوادگی و یک مجموعه داده ی تلفن همراه دنیای حقیقی، یک مجموعه داده ی تلفن همراه شبیه سازی شده را ایجاد کردند که دارای اطلاعات درستی زمین است و الگوهای مکانی زمانی مشابه را به صورت مجموعه داده ی تلفن همراه شبیه سازی شده نشان میدهند. این امر سبب میشود آنها بتوانند صحت نتایج استنباط شده را در سطح انفرادی ارزیابی نمایند. یک رویکرد مبتنی بر مدل هم برای شناسایی موقعیتهای مکانی فعالیت و استنباط نوع آنها توسعه پیدا کرد. نتایج کاملا نوید بخش است- تعداد متوسط موقعیتهای منحصر به فرد بازدید شده با رویکرد مبتنی بر مدل، 2.73 است در مقایسه با 2.82 به صورت درستی زمین و 4.62 با استفاده از رویکرد متداول در ادبیات. به علاوه، 70٪ و 65٪ از مکانهای منزل و محل کار شناسایی شده، در 100 متری از موارد صحیح قرار دارند. روشهای دیگر هم مورد بررسی قرار خواهند گرفت. برای نمونه، محققان حمل و نقل، از دیرباز هم ردیابی های جی پی اس و هم داده های زمینه یابی را جمع آوری کرده اند، از داده های زمینه یابی به صورت درستی زمین برای اهداف اعتبار یابی استفاده میکنند. روشهای مشابه میتوانند به منظور اعتبار یابی نتایج استنباط شده از داده های بزرگ پذیرفته شوند، که با رواج دستگاه های موبایل در بسیاری از جمعیتها تسهیل شده است.

2.3.2 معرف بودن

یک مساله ی اصلی که به طور وسیعی در تحقیقات داده های بزرگ کنونی نادیده گرفته شده است، معرف بودن داده هاست. به طور وسیعی تصدیق شده است داده های تلفن همراه، معرف نیستند و دلایل متعددی در این مساله دخیل است من جمله نزدیکی تلفنهای موبایل، تلفنهای موبایل متعدد، سرعتهای نفوذ که به شدت با حاملین تلفن همراه تغییر می یابند و انتخاب نمونه. از آنجاییکه فراوانی استفاده از تلفن به شدت بین کاربران فرق دارد، کاربران فعالتر، احتمالا در این نمونه نشان داده میشوند و آن دسته از کاربرانی که به طور متحرک تر از بقیه یافته شده اند. لذا، انتخاب نمونه بر مبنای استفاده ی تلفن، به طور بالقوه به بیش برآورد مقادیر تحرک منجر خواهد شد. از سوی دیگر، برخی مطالعات هم پیشنهاد شده است که به نظر میرسد برخی سنجشهای تحرک برای این سوگیری نمونه برداری، مصون هستند. آنگاه سوال این طور میشود که چه سنجشهای تحرکی

تمایل دارند نسبت به این سوگیری مصون باشند، تحت چه شرایطی و چه اطلاعاتی میتواند از این سنجش های تحرک استنباط شود.

برای داده های بزرگی که در برنامه ریزی حمل و نقل برای تصمیمات مربوط به سرمایه گذاری و خط مشی بکار میروند، مساله ی معرف بودن باید مورد بررسی قرار گیرد. اتصال الگوهای تحرک به مطالعه جامع شناسی- جمعیت شناسی، حائز اهمیت است، نه تنها در سطح ناحیه ای (مثلا شی و همکارانش، 2015)، بلکه مهم تر از آن در سطح انفرادی. یک مطالعه ی جدید در مورد پیش بینی وضعیت ثروت و فقر در سطح انفرادی با استفاده از داده های CDR همان فرد، نتایج نوید بخشی را نشان میدهد، که حاکی از این است سوگیری جمعیت وابسته به داده های تلفن همراه ممکن است به طور بالقوه، اصلاح شود. همانطور که از بخش 2 در مورد ماهیت داده ها به بخشهای 3-5 در مورد بخشهای فرعی مهم در تحلیلهای رفتار سفر (تحرک انسان) گذر میکنیم، در ذهن نگه داشتن این مسائل مربوط به داده، حائز اهمیت میشود، زیرا هم در سوالات و هم در پاسخ ها دخیل هستند.

3. عوامل رفتاری

یک چالش در شناخت و پیش بینی الگوهای تحرک انسان، توضیح عواملی است که موجب الگوهای رفتاری مشاهده شده میشود. دهه های متمادی تحقیقات رفتار سفر ممکن است در این زمینه روشنگری زیادی ایجاد نماید. تقریبا برای همه گونه های رفتار سفر، سه دسته عامل همیشه مهم یافته شده اند: جامع شناسی- جمعیت شناسی، محیط ساخته شده (مثلا تاثیر تراکم، تنوع، طراحی و کاهش مسافت) (مثلا اثر لی و همکارانش، 2014، پیرامون اثر کاهش مسافت) و سفر و عوامل مربوطه ی دیگر (مثلا هدف سفر، زمان سفر و هزینه ها، قابلیت اطمینان). الگوهای تحرک مختلف شناسایی شده از داده های بزرگ به طور بالقوه استراتژی های جستجوی مختلفی را نشان میدهد که توسط افراد بکار گرفته میشوند- برای نمونه، پرواز لوی ناقص (TLF) شناسایی شده، یک الگوی جستجوی تحت تاثیر محیط ساخته شده را پیشنهاد میدهد، تاییداتی برای میزان وسیعی از مطالعات ارائه میدهد که نقش محیط ساخته شده را بر رفتارهای سفر بررسی میکند (مثلا چن و همکارانش، 2008). از آن مهم تر اینکه، مطالعات جدید در مورد اینکه الگوهای تحرک ممکن است چطور تحت محیطهای ساخت شده ی مختلف تغییر کنند، در کمک به ما برای درک شیوه ای که استراتژیهای جستجوی افراد ممکن است تحت شرایط مختلف تغییر کنند، مفید خواهد بود. در واقع، یافتن الگوهای تحرک مختلف (مثلا TLF،

نمایی و غیره) از تحقیقات داده ای بزرگ اخیر حاکی از این است حرکت انسان، به سادگی حاصل یک فرایند مکانیکی نیست، بلکه تا حد زیادی تحت تاثیر شماری از عوامل است که از مکانی به مکانی و از جمعیتی به جمعیتی دیگر فرق میکنند.

تهییج در این زمینه فقط بر این نیست که چه عواملی بر تصمیمات اثر دارند بلکه بیشتر بر شیوه ای است که این عوامل در پیچیدگی های اساسی مشمول در تصمیمات سفر روشنگری نموده اند. در زبان آمار، در مورد آشکار نمودن مکانیسم های علیت اساسی است. در تحقیقات حمل و نقل، تئوری های دیگری (از تئوری به حداکثر رسانیدن مطلوبیت تصادفی) محک زده شده اند به عنوان مثال، تئوری آینده نگری، منطق کران دار به عنوان بخشی از فرایندهای تصمیم گیری و تئوری تاسف. در شرایط وسیعتر و طولانی تر تحلیل، تئوری رفتار برنامه ریزی شده ی سفر، تئوری فعالسازی-استاندارد و ترکیبی از هر دو به عنوان راه دیگری برای بررسی بکار رفته اند و رفتارهای انتخاب انفرادی را شرح میدهند. به طور همزمان، مطالعات جدیدتر با استفاده از داده های بزرگ برای تحقیق در مورد مکانیسم های اساسی آغاز شده اند که به الگوهای مشاهده شده منجر شده اند. در عین حال، این مطالعات به طور وسیعی در تلاش هستند الگوهای امضا(فیزیکی) را شناسایی نمایند که با هم الگوهای تحریک مشاهده شده را میسازند. این نتایج (مثلا اشکال عمده ی تحرک انسانی، الگوهای بازگشت ترجیحی)، بیتش آفرین هستند، اما مکانیسم های اساسی برای انتخابهای سفر انفرادی را شرح نمیدهند. با این حال، از سوی دیگر، بر مبنای این الگوهای امضای شناسایی شده، میتوانیم تئوری ها و فرضیاتی را گسترش دهیم درباره ای اینکه افراد چگونه انتخابهای مربوط به سفر خود را انجام میدهند و مدلهای شبیه سازی را برای تایید این فرضیات بکار میبریم. به عنوان مثال، میتوان این سوال را پرسید "آیا الگوهای امضای شناسایی شده، نتایج هیچ یک از تئوریهای رفتاری فوق هست یا خیر (مثلا منطق کراندار، تئوری آینده نگری) یا شاید ترکیبی از آنها؟". جوابها به این سوال، نه تنها شناخت ما در رفتارهای سفر افراد بلکه قابلیت‌های پیشگویانه در پیش بینی تقاضای سفر را گسترش میدهد. و چنین تحقیقاتی به همکاری از سوی هم محققان داده های بزرگ و هم داده های کوچک نیاز دارد.

از آنجاییکه یک سفر واحد، با بیشترین احتمال در اتصال به سفرهای دیگر صورت گرفته ی قبلی یا بعدی انجام میشود، زنجیر کردن سفر از دیر باز مورد تاکید بسیاری از مطالعات بوده است، چه از یک چشم انداز روزانه دیده

شود (مثلا فعالیت روزانه و الگوهای سفر) یا از یک سطح تور (مثلا دنباله ای از سفرهایی که در یک موقعیت مکانی شروع و تمام میشوند). از اینرو، پیشرفت فعالی در مدل‌های مبتنی بر فعالیت وجود داشته است که سعی دارد اتصالات بین سفرها را در یک روز شرح دهد. چون سفرها در بافتی روی میدهند که فرد زندگی میکند و کار میکند، تحقیقات فعال بسیاری در شناخت و مدلسازی اتصال بین انتخابهای دراز مدت (مثلا انتخاب موقعیت محلی) و رفتارهای سفر سازی وجود داشته است. یک موضوع که توجه زیادی را در دهه ی گذشته جلب نموده است، خود گزینی محلی است- افراد خودشان محله های دارای ویژگی هایی را انتخاب میکنند که با سبک زندگی شان تطبیق دارد. به خاطر وجود خود گزینی محلی، مطالعات ارزیابی اثر محیط ساخته شده در مورد رفتار سفر، باید خود گزینی محلی را کنترل نمایند. شماری از مطالعات هم در چشم انداز چرخه ی حیات انجام شده اند، که در صدد شناخت ریشه ی خود گزینی محلی است. به خاطر ماهیت طولی، داده های بزرگ میتوانند در این حوزه ی خاص روشنگری ایجاد نمایند با شناسایی افرادی که محل زندگی یا کار خود را جابجا کرده اند و تحلیل میکند الگوهای تحرک چطور ممکن است تغییر یافته باشند. برای رها کردن خود گزینی محلی (از رفتار سفر به محیط ساخته شده) از اثر محیط ساخته شده (از محیط ساخته شده به رفتار سفر)، مدل‌های آماری پیشرفته، که محققان رفتار سفر با آن آشنا هستند (مثل مدل‌های معادلات ساختاری) میتواند بکار گرفته شود.

تحقیقات در زمینه ی کاوش داده های بزرگ میتواند به شناسایی عوامل اضافی کمک کند که تصمیمات سفر ما را شرح میدهند و مکانیسم های تصمیم زیربنایی را آشکار میکند. به عنوان مثال، شماری از مطالعات اتصال بین شبکه های اجتماعی فرد و رفتار سفر را هم در زمان و هم در مکان نشان میدهند. با فرض این مساله، ممکن است بسنجیم آیا روابط اجتماعی فرد به رفتارهای سفر متعاقب منتهی خواهد شد یا بالعکس؟ شاید روابط در هر دو شیوه وجود داشته باشد. جوابها به این سوالات، چندین مضمون مهم دارند من جمله در عوامل اضافی روشنگری تازه ای ایجاد میکنند که موجب رفتارهای مشاهده شده میشود، که فرصتهای مداخله ی احتمالی را برای اصلاح رفتار سفر از طریق شبکه ی اجتماعی او ارائه میدهد و تعاملات اجتماعی را از طریق تغییرات رفتار سفر بر می انگیزاند و تغییر میدهد.

پیشرفت‌ها در تحقیقات رفتار سفر هم چنین می‌توانند به طور همزمان تحقیقات داده‌های بزرگ را در تحلیل الگوهای تحرک انسانی پیش ببرند. تحقیقات رفتار سفر، به عنوان مثال از دیر باز نشان داده است نگرشها و اولویتهای یک شخص می‌تواند در انتخابهای سفر، نقش مهمی را ایفا نماید. و بر خلاف عقیده دیرینه ای که نگرشها در طول زمان پایدار می‌مانند، تحقیقات کنونی نشان دادند نگرشهای فرد می‌تواند پس از تغییر رفتار، تغییر نماید. شاید این نتایج بتواند تحقیقات داده‌های بزرگ را به وضعیت فراتر از وضعیت کنونی سوق دهد: در اصل با تحلیل الگوهای حرکت مشاهده شده و با نادیده گرفتن تجارب ایجاد و نگرشهای شکل گرفته از طریق این حرکات. در تحقیقات رفتار سفر، نگرشها اغلب از مجموعه‌ای از گفته‌های نوشته شده توسط محققان بیرون کشیده میشوند، در حالیکه اشکال طبیعی از طریق داده‌های بزرگ (مثلا داده‌های رسانه‌ی اجتماعی تیوتر) ممکن است مشاهداتی جدیدی را در مورد نگرشها و ارزشها آشکار نماید و توسعه‌ی متدولوژیکی را تحریک نماید (مثلا پردازش زبان طبیعی). به علاوه، یک مرحله‌ی فراتر از شناخت تقریبی همبستگی‌ها، اما نسبت به افشای مکانیسم‌های علیت، مورد استقبال خواهد بود. در این زمینه، طراحی‌های تحقیق (مثلا طراحی‌های شبه آزمایشی) و روشها (مثلا سیستم‌های معادلات ساختاری) به طور وسیعی در تحقیقات رفتار سفر برای باز کردن علیت بکار می‌روند که ممکن است گسترش پیدا کند.

4. توسعه مدل

در تحقیقات حمل و نقل، مدل‌های انتخاب گسسته، مدل اساسی توسعه یافته برای پیش بینی رفتارهای سفر، انتخاب‌های ویژه‌تر در مورد حالات حمل و نقل، زمان روز، مقاصد و حتی مسیرها هستند. مدل‌های انتخاب گسسته بر مبنای نظریه‌ی حداکثر سازی مطلوبیت تصادفی (RUM) توسعه می‌یابند. RUM فرض میکند زمانیکه از یک مجموعه از گزینه‌های گسسته انتخاب صورت می‌گیرد (مثلا انتخاب‌های حالت و انتخابهای مقصد)، تصمیم گیرنده از کل گزینه‌های امکان پذیر و صفات وابسته‌ی آنها آگاه است، خواهان ایجاد تبادلاتی بین صفات و با فرض این شروط است، آنگاه گزینه‌ای را انتخاب میکند که رضایتش را به حداکثر خواهد رساند. هم شواهد حکایتی و هم علمی حاکی از این است این فرضیات فقط برای مجموعه‌ی خاصی از رفتارهای تحت شرایط خاص بکار می‌رود. مدل‌های انتخاب گسسته‌ی استاندارد مثلا مدل لاجیت چند جمله‌ای، معمولا برخی فرضیات مشترک را حفظ میکنند، یعنی: 1) مولفه‌های تصادفی مطلوبیت گزینه‌های مختلف، مستقل هستند و

به طور یکسان (IID) با توزیع ارزش نهایی نوع 1 (یا گامبل) توزیع میشوند؛ 2) همگنی را در پاسخدهی به صفات گزینه ها بین کل افراد حفظ میکنند (یعنی فرضیه همگنی پاسخ)؛ و 3) ساختار واریانس- کوواریانس خطای گزینه ها، بین افراد مشابه است (یعنی یک فرضیه همگنی واریانس- کوواریانس خطا). در عین حال، این فرضیات نقض میشوند چنانچه مسافریین مطلوبیت بالاتری را برای کل حالات ترانزیت (اتوبوس، قطار و غیره) دریافت نمایند به خاطر فرصت ایجاد تعادل در قطار، هواپیما و ... با مسافریین دیگر یا چنانچه سیستم حمل و نقل عمومی سطوح راحتی مختلفی را در مسیرهای مختلف (یک متغیر نامشهود) ارائه دهد. برای بررسی این مساله، سه دسته از مدل‌های انتخاب گسسته که یک یا بیش از یکی از فرضیات مورد بحث در بالا را تضعیف میکنند اکنون به طور عادی بکار رفته اند. دسته ی اول مدلها (به صورت "مدل‌های واریانس ناهماهنگ" برچسب گذاری شده اند) نسبتاً محدود کننده است: آنها فرضیه ضریب خطای توزیع شده به طور مشابه (را در بین کل گزینه ها) تضعیف میکنند اما فرضیه ی استقلال را (بخشی از فرضیه اول فوق) کاهش نمیدهند یا فرضیه همگنی پاسخ (فرضیه دوم فوق). دومین دسته ی مدلها (که به صورت "مدل‌های لاجیت چند جمله ترکیبی (MMNL) برچسب گذاری میشود) و دسته ی سوم مدلها (که به صورت "مدل‌های مقدار نهایی تعمیم یافته ی ترکیبی (MGEV)" برچسب گذاری میشود) بسیار کلی هستند؛ مدل‌های موجود در این دسته، برای کاهش ساختار خطای استقلال و توزیع شده به طور مشابه (سرتاسر گزینه ها) در MNL هم چنین کاهش فرضیات همگنی پاسخ، به قدر کافی انعطاف پذیر هستند. برای بحث بیشتر در این زمینه، خواهشمند است اثر بت (2000) را ملاحظه نمایید. کشف الگوهای تحرک خاص از داده های بزرگ، فرصتی را برای شناسایی پیوندهای بین انتخاب های انفرادی میکروسکوپی و رفتارهای ماکروسکوپی پدید آمده به ما ارائه میدهد و قواعد تصمیم گیری بکار رفته برای مدلسازی انتخابهای مربوط به سفر را مجدداً ارزیابی میکند. یک سوال نخست حتمی که میتوان پرسید این است: اگر هر فردی تصمیمات مربوط به سفر بگیرد همانطور که از طریق نظریه مطلوبیت تصادفی بکار رفته به طور وسیع مشخص شده است و در نتیجه مدل‌های انتخاب گسسته، چه نوع الگوهای تحرکی در سطح جمعیت پدید خواهد آمد؟ چه چارچوبهای ریاضیاتی، الگوهایی را ایجاد خواهند کرد که به طور دقیق تری از چارچوب مشاهده شده در سطح جمعیت تقلید میکنند؟ شبیه سازی های مبتنی بر عامل ممکن است برای پاسخ دهی به این سوالات بکار روند.

به علاوه، یک مدل اساسی وجود ندارد (مثل مدل‌های انتخاب گسسته) تا شیوه ای را پیش بینی نماید که رفتارهای سفر انفرادی ممکن است در طول زمان تغییر یافته باشند. در این زمینه، توسعه در مطالعات داده های بزرگ اخیر میتواند مشاهداتی را ارائه نماید. سانگ و همکارانش (2010) نشان دادند فراوانی بازدید k مرتبه ای یک فرد از موقعیت مکانی دارای بیشترین بازدید f_k میتواند به صورت $f_k \sim k^{-2}$ تخمین زده شود، که در آن $1.2 \pm 0.1 \approx \xi$. هم چنین بیشترین قابلیت پیش بینی را محاسبه نمودند که احتمالی است که یک الگوریتم پیش بینی مناسب میتواند به درستی جای تقریبی کاربر، پیکها را در حدود 0.93 پیش بینی نماید. نتایج مشابه از طریق محققان دیگر یافته شده اند.

حوزه ی پیشرفت بالقوه ی دیگری به فرمولاسیون مجموعه های انتخاب برای مدل‌های انتخاب مقصد مربوط است. این یک قرارداد بوده است که مجموعه ی انتخاب برای مدل‌های انتخاب گسسته شامل کل موقعیتهای مکانی مختلف در یک ناحیه میشود. این شیوه برای مدلسازی هم انتخابهای موقعیت اقامتی طولانی مدت و هم انتخابهای کوتاه مدت برای فعالیتهای تفریحی و خرید بکار رفته است. به طور وسیعی به رسمیت شناخته شده است چنین رویکرد فراگیری، غیر واقع گرایانه است و احتمالاً با شیوه ای که افراد در دنیای واقعی جستجو میکنند، مطابقت ندارد. مطالعات در تحقیقات حمل و نقل سعی داشته اند، چارچوب انتخاب گسسته را هم چنین برای مدلسازی فرایند ایجاد مجموعه ی انتخاب برای شرح برخی عوامل (مثلاً نگرشها، ادراک و معیارهای حذف) گسترش دهند. اینجا، نتایج حاصل از تحقیقات داده های بزرگ ممکن است توزیع ویژه ای را ارائه نماید که انتخابهای مختلف بتوانند از آن ایجاد شوند. به عنوان یک مثال، برخی از مطالعات داده های بزرگ، یک الگوی پرواز لوی ناقص را برای تحرک انسان نشان داده اند زمانیکه جابجایی (بین برجهای سلولی) بکار میرود، حاکی از این است حرکات مقیاس کوچک، به طور تصادفی تری توزیع میشوند (بیشتر یک الگوی دایره ای) در حالیکه حرکات مقیاس بزرگ بیشتر احتمال دارد یک تابع توزیع فضایی فرصتها در آن ناحیه باشند (بیشتر مسیرهای پشت سر هم جهت دار). این امر بدین معنی است تشکیل مجموعه های انتخاب احتمالاً فرق میکند که به نوع انتخابهای مورد نظر بستگی دارد- شاید رویکرد جامع موجود برای انتخابهای کوتاه مدت، مناسب تر باشد همانند فعالیتهای خرید و تفریحی اجتماعی و برای انتخابهای دراز مدت همانند موقعیتهای مکانی مسکونی و محل کار، آنچه در یک ناحیه وجود دارد ممکن است نقش مهم تری را ایفا نماید. این مسائل هنوز هم بدون پاسخ می ماند.

5. الگوهای تحرک انسانی

در اصل به خاطر کمبود داده های طولی، اکثریت مطالعات قبلی در تحقیقات رفتار سفر از داده های سطح مقطعی استفاده میکنند. شمار محدودی از مطالعات، داده های کوچک را بر روی یک دوره ی مدید جمع آوری نموده اند از سه تا چهار روز گرفته تا بیش از یک ماه و طولانی تر. نمونه های اخیر شامل زمینه یابی پنل خانوار بریتانیایی، که اکنون در بررسی طولی خانوار انگلستان تلفیق شده است و چندین مجموعه داده ی دیگر که در آن برای اهداف ویژه ای جمع آوری شدند. متریکهای بکار رفته شامل نرخ های سفر روزانه، زمان سفر روزانه، مسافت سفر روزانه یا ترکیبی از بخشی یا کل فضای عملکرده، مصرف زمان فعالیت و دنباله های سفر "منحصر به فرد" میشود. این مطالعات هم میزان قابل توجهی از تکرار و هم تغییر پذیری را آشکار مینمایند زمانیکه بین روزهای متوالی مقایسه میشود، یا در ترکیبات دو صفت ارزیابی میشود (مثلا خرید و خودرو). میزان تغییر پذیری درون فردی از بیش از 20٪ تا حدود 80٪ متفاوت یافت میشود که به متریک های بکار رفته، نوع سفرها و هم چنین روزهای تحلیل بستگی دارد. مطالعاتی که شاخصهای شباهت را توسعه دادند، ابعاد متعدد یک فعالیت روزانه را ترکیب مینمودند و هم چنین الگوهای سفر به نتیجه گیری یکسانی دست می یابند.

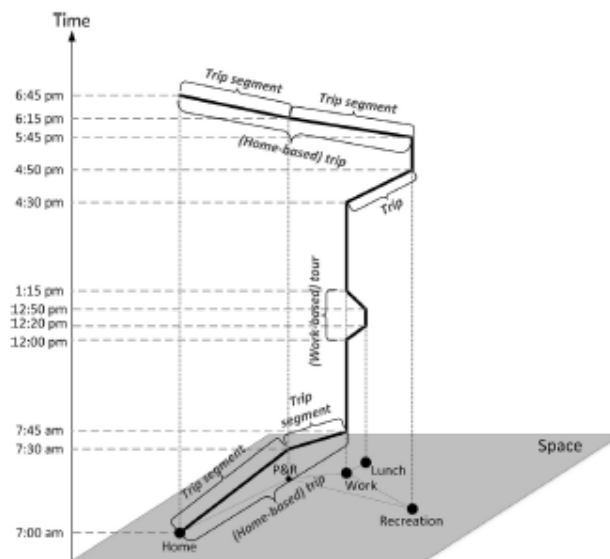
مطالعات با استفاده از داده های بزرگ گرایش به تمرکز بر شناسایی الگوهای تحرکی دارند که از آنها میتوان پیش بینی را انجام داد. به جای استفاده از متریک های معنادار مثل نرخ های سفر، مسافت سفر یا زمان سفر، این مطالعات، جابجایی های بین توقف های استنباط شده (مثلا بین دو موقعیت برق سلولی) را تحلیل میکنند و آنها را به صورت دنباله هایی از رویدادهای تصادفی در نظر میگیرند که فرض میشود توزیعات فضایی و زمانی شان دارای نظمهای آماری خاصی باشد. این نتایج در حالیکه آگاهی بخش هستند، با استفاده از متریکهایی (مثلا جابجاییها) استنتاج میشوند که مفهوم شان واضح نیست. در اکثر این مجموعه داده ها که شامل برجهای تلفن همراه میشود، جابجایی ها، مسافتهای خط راست بین دو برج تلفن همراه مشاهده شده به طور متوالی هستند. یک نکته ی اضافی این است که سنجشهای مسافتهای خط راست، از مسافتهای شبکه ای، کمتر معنادارند زیرا هم انسانها و هم وسایل نقلیه، راه های تعریف شده ای را بر روی زمین، تحت رایج ترین شرایط، دنبال میکنند. در مجموعه داده های شامل مشاهدات تلفن همراه، مفهوم جابجایی، وضوح بیشتری دارد، همانطور که باید اولین بار موقعیتهای مکانی فعالیت را (یا مبداها و مقصدها) از داده ها استنباط نمود. با این حال از آنجاییکه اکثر

مطالعات سعی داشته اند موقعیتهای فعالیت را به جای خط سیرها مورد کاوش قرار دهند و مطالعات معدودی سعی کرده اند آن را تایید کنند، مفهوم یک جابجایی هنوز هم نامشخص است. به بیان دیگر، استفاده ی مستقیم جابجایی ها در کاربردهای حمل و نقل، با پیامدهای ناشناخته، نامناسب است.

یک دستورالعمل نخست فعلی برای محققان داده های بزرگ، وضوح مفهوم یک جابجایی است و این امر مستلزم دو مرحله ی ضروری است: اولاً، اعتباریابی روشهای بکار رفته برای کاوش موقعیتهای فعالیت و خط سیرها از داده های بزرگ و ثانیاً، استفاده از شرایط معنادار همانند بخشها، سفرها، تورها یا الگوها. همانطور که پیشتر ذکر شد، میزان محدود کارهای صورت گرفته بر روی اعتباریابی، تقریباً در سطح ناحیه ای است، به عنوان مثال مقایسه ی تراکم جمعیت از داده های بزرگ در برابر داده های سرشماری محاسبه شد یا مقایسه ی سهم های اهداف سفر در برابر زمینه یابی سفر. در حالیکه این تلاشهای اولیه، نوید بخش هستند، این رشته به مطالعات بسیار بیشتری نیاز دارد به طور خاص، مطالعاتی که بتوانند این اعتبار یابی را در سطح انفرادی انجام دهند، به طور صریح به سوالاتی چون "چه درصدی از افراد در یک نمونه، موقعیتهای مکانی و خط سیرهای فعالیتشان به طور صحیح یا غلط استنباط شده است؟" پاسخ دهند.

برای شناخت متریکهای مربوط به سفر معنادار مکرراً بکار رفته توسط محققان حمل و نقل، شکل 5 یک نمایش سه بعدی از فرضیه منطقی *what is* برای یک محقق حمل و نقل، یک بخش، یک سفر، یک تور و یک الگوی روزانه را ارائه میدهد. یک سفر به حرکت بین دو موقعیت مکانی معنادار اشاره دارد، به عنوان مثال از خانه به محل کار در صبح، که هم چنین یک سفر مبتنی بر خانه است (به صورت سفرهایی تعریف میشود که یک سر آن، خانه است). در شکل 5، سفر از خانه به محل کار دارای دو بخش است، که از یک رانندگی به یک ناحیه P&R و یک اتوبوس سواری (رسم شده با دو شیب برای سرعتهای سفر مختلف) تشکیل میشود. در ظهر، شخص از محل کار به یک محل ناهارخوری میرود، در آنجا ناهار میخورد و بعد به محل کار بر میگردد. این دنباله ی دو سفره (از محل کار به ناهار و بعد برگشت به محل کار) یک توری مبتنی بر محل کار را تشکیل میدهد (که به صورت یک دنباله از سفرها تعریف میشود) که در یک موقعیت مکانی شروع شده و تمام میشود). هر دوی این سفرها، سفرهای مبتنی بر غیر خانه هستند. پس از کار، این شخص به یک محل تفریحی میرود، در آنجا مدت زمانی را سپری میکند و سپس به خانه بر میگردد. سفر از محل کار به تفریح، بار دیگر مبتنی بر غیر خانه است،

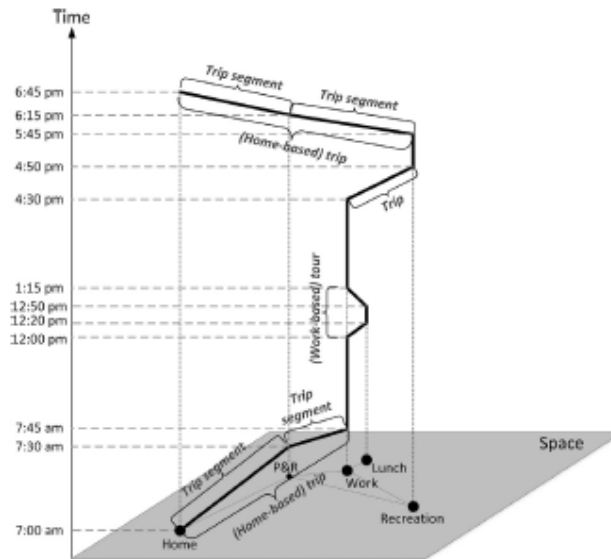
اما سفر از تفریحگاه به خانه، مبتنی بر خانه است. روی هم رفته، در کل سفرها، از سفر اول، فرد خانه را ترک میکند تا سفر آخر، که فرد به خانه بر میگردد، از یک فعالیت روزانه و یک الگوی سفر تشکیل میشود.



شکل 5. یک نمایش سه بعدی از سفرها، بخشهای سفر، تورها و فعالیت روزانه و الگوهای سفر.

به علاوه، هر زمان شبکه های حمل و نقل دنیای واقعی امکان پذیر، به تصویر کشیده خواهند شد، به این ترتیب مسافتهای شبکه ی معنادار، سنجیده میشوند. زمانیکه این مراحل تکمیل شوند، بسیاری از سوالات پیگیری مهم ممکن است مطرح شوند از جمله به عنوان مثال:

1. زمانیکه با مبداها و مقصدهای آشکار و در متریکهای معنادار (مثل زمان سفر، مسافتهای سفر/ شبکه) سنجیده میشود، چه الگوی تحرک خاصی از داده های بزرگ پدید خواهد آمد؟
 2. یک الگوی مشاهده شده چطور تغییر می یابد زمانیکه حالات مختلف حمل و نقل یا اهداف سفر مختلف مورد بررسی قرار میگیرند؟
 3. نقش محیط ساخته شده ی اساسی در تعدیل الگوهای مشاهده شده چیست؟
- دیدگاه ما این است که پتانسیل کامل ماهیت طولی داده های بزرگ (اغلب بیش از 1 ماه به طول می انجامد) هنوز هم کاملاً مورد بررسی قرار میگیرد. این مساله به طور خاصی حائز اهمیت است زمانیکه متریکهای معنادار بکار میروند. ما برخی سوالات نمونه را مطرح میکنیم و



شکل 5. یک نمایش سه بعدی سفرها، بخشهای سفر، تورها و فعالیت روزانه و الگوهای سفر.

به هیچ وجه این سوالات مجموعه ی کاملی از سوالات را در بر نمیگیرند که به طور قابل توجهی دانش ما را در الگوهای تحرک انسانی بالا ببرند:

1. ریتم های گوناگون نظم (مثلا روزانه در برابر هفتگی) با انواع مختلف سفر چیست و این چرخه ها چگونه با محیط ساخته شده تحت تاثیر قرار میگیرند؟

2. این ریتم های گوناگون چگونه با همدیگر در تشکیل الگوهای تحرک فرد در طول زمان تعامل برقرار میکنند؟

3. تغییرپذیری درون فردی، چگونه در طول زمان تغییر میکند؟

6. بحثها

این مقاله با یک هدف ساده ی تلفیق دو گروه از محققان در تحلیل الگوهای تحرک انسان، به رشته ی تحریر در آمده است: محققان رفتار سفری که از دیر باز بر زمینه یابی های سفر خانواده (داده های کوچک) تاکید داشته اند و محققان داده های بزرگ که اخیرا از داده های ایجاد شده به طور انفعالی (داده های بزرگ) استفاده نموده اند. بدین منظور، برخی مفاهیم و پیشرفتهای مهم را در هر رشته معرفی میکنیم و برخی ایده های میان رشته ای را در سه ناحیه فرعی مطرح می نماییم: (1) عوامل رفتاری (بخش 3؛ 2) توسعه مدل (بخش 4؛ و 3) الگوهای حرکت (بخش 5).

فهرستی از ایده های مطرح شده، به هیچ وجه یک فهرست جامع نیست، بلکه هدف مهم برانگیختن یک بحث را از هر دو گروه محققان تامین مینماید و همکاری هایی را بین دو گروه ایجاد مینماید. در واقع، مقالات اخیر یک همگرایی بین دو گروه را نشان داده اند. برخی از تئوری های معروف در رفتار سفر، در تحقیقات داده های بزرگ مجدد کشف شده اند، برای نمونه، پایداری های فضایی و زمانی وابسته به فعالیتهای مختلف به خاطر محدودیتهای گوناگون (مثلا زیستی)، پیشینه و زمان وابستگی روز و اهمیت شناخت فعالیتهای پشت سر سفرها. این کشف های مجدد، مورد استقبال قرار میگیرند، اما این رشته با داشتن ارتباطات بیشتر بین گروه ها، بسیار بیشتر مفید خواهد بود. به علاوه، مطالعات داده های بزرگ، حرکت به سوی یک مقیاس تحلیل را در سطح انفرادی آغاز نموده اند، که تایید میکند احتمالا میزان قابل توجهی همگنی در یک جمعیت وجود دارد، که پیشینه ی دراز مدت چندین دهه ای را در تحقیقات مبتنی بر فعالیت مبتنی بر فرد را با محققان حمل و نقل تایید میکند. این مقاله احتمالا به ناچار نواحی اضافی را صرف نظر کرده است که همکاری های بین دو گروه به طور دوجانبه مفید خواهد بود. یک ناحیه، تعیین بخشها در یک ناحیه مطالعاتی است. در حالیکه محققان حمل و نقل از دیرباز از بخشهای تحلیل ترافیک (TAZ) استفاده کرده اند، محققان داده های بزرگ اغلب بر گریدها یا شبکه های یکنواخت متکی هستند. گفتگوهای بین دو گروه میتواند به طور بالقوه ای به برخی سوالات اساسی پاسخ دهد همانند اینکه چه چیزی یک بخش را در زمینه ی تحلیل الگوهای تحرک انسانی تشکیل میدهد و چه روشهایی را میتوان برای شناسایی یک بخش با استفاده از داده هایی بکار برد که در دسترس هستند؟ مزایای زیادی را میتوان استنتاج نمود چنانچه محققان حمل و نقل و داده های بزرگ بتوانند سعی کنند به این سوالات با هم پاسخ دهند. همکاری های حاصل بین این دو ممکن است به مطالعات کیفیت بالایی منجر شود که تحلیل هایشان بر مبنای شناخت صحیح شیوه ای است که افراد در زمان و فضا حرکت میکنند. چنین مطالعاتی به همان اندازه مستقیما به توسعه ی تصمیمات و خط مشی های حمل و نقل مرتبط خواهد بود، به گونه ای که چطور میتوانیم خط مشی هایی (کوتاه مدت و دراز مدت) را گسترش دهیم تا رفتارهای سفر مساعدتر برای محیط زیست را تشویق کنیم و آینده ای پایدارتر را بسازیم. یک نکته ی احتیاط دیگر همانطور که به سرعت به سوی تحقیقات داده های بزرگ سوق می یابیم، آگاهی از میزان سفسطه هایی است که در طول این مسیر خلق خواهیم کرد. به بیان میشل جردن، قیاس، پیش بینی چیزی با میزان انبوهی از داده ها به این صورت است

"داشتن میلیاردها میمون که تایپ میکنند و یکی از آن شکسپییر را خواهد نوشت". این دیدگاه ماست که مفهوم سازی چارچوبها و فرضیات ایده ها، همانطور که از دیرباز توسط محققان حمل و نقل با داده کوچک بکار رفته است، رها نخواهد شد و کاربردهای مداوم تکنیکهای تحقیقاتی مهم، فقط تحقیقات داده های بزرگ را تقویت خواهد کرد. تردیدی وجود ندارد و به طور همزمان، آگاه بودن از خطراتشان، به همان اندازه مهم است. در واقع این خطرات است که همکاری ها را با محققان حمل و نقل به طور بالقوه متقاطع میکند.

تقدیرنامه ها

منابع

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