گزارش چالش برآورد ماتریس مبدا و مقصد مسافران اتوبوس و مترو

سید معین امیری*[۱]، فاطمه شامحمدی [۲]، علیرضا شکرآرا [۳]، شمیم عباسی[۴]، امیرسجاد صادقی[۵]

[۱] گروه AID، s.moeinamiri@gmail.com

[۲] گروه AID، fatemeshm@gmail.com

[۳] گروه AID، alrzsh٤٨@gmail.com

[۴] گروه AID، shmimabbasi٧٩@gmail.com

[۵] گروه AID، sadeqi.amirsajad@gmail.com

چکیده - هدف از پژوهش حاضر برآورد ماتریس مبدا-مقصد مسافران حمل‌ونقل همگانی با استفاده از داده‌های سامانه‌های هوشمند پرداخت کرایه است. با توجه به در دسترس نبودن داده‌های *AVL* اتوبوس‌ها به‌عنوان داده‌های کلیدی مورداستفاده در شناسایی و تشخیص ایستگاه‌ها، چندین روش مختلف جهت شناسایی ایستگاه‌های سوار و پیاده شدن مورداستفاده و بررسی قرار گرفت. در الگوریتم طراحی‌شده ایستگاه‌های سوار و پیاده شدن با در نظرگیری توالی سفرها و با استفاده از اطلاعات قطعی موجود (ایستگاه‌های سوارشدن مترو و *BRT*) و همچنین تلاقی خطوط با یکدیگر برآورد گردید. نتایج نشان می‌دهد که با یک مرحله اجرای الگوریتم، ایستگاه‌های بیش از ۷۳ درصد از سفرها حاصل‌شده که در صورت تکرار آن میزان شناسایی افزایش قابل‌توجهی خواهد داشت. درنهایت میزان مسافر هر ناحیه نیز پس از تخصیص مسافران هر ایستگاه به نواحی مربوطه حاصل گردید.

کلیدواژه- ماتریس مبدا-مقصد، سامانه هوشمند پرداخت کرایه، توالی و تلاقی سفرها

۱- مقدمه

امروزه سامانه‌های هوشمند پرداخت کرایه[۱] به‌عنوان روشی برای جمع‌آوری کرایه حمل‌ونقل همگانی مورداستفاده قرار گرفته است. این سامانه‌ها در کنار دریافت کرایه به‌طور خودکار و پیوسته سوابق استفاده مسافران از وسایل حمل‌ونقل عمومی را نیز با ثبت اطلاعات شناسایی کارت هوشمند مسافران جمع‌آوری می‌کنند. تجزیه‌وتحلیل این اطلاعات می‌تواند ضمن ارائه مواردی نظیر حجم تقاضای هر ایستگاه و خط، تعداد دفعات سفر هر مسافر، توالی سفرها و سفرهای چندگونه‌ای، در مدیریت و برنامه‌ریزی حمل‌ونقل مورداستفاده قرارگرفته و منجر به بهبود کیفیت خدمات ارائه‌شده شود. ازاین‌رو هدف این پژوهش تخمین ماتریس مبدا-مقصد (که بیانگر میزان تقاضای سفر از هر ناحیه به ناحیه دیگر است) با استفاده از بررسی و داده‌کاوی داده‌های سامانه هوشمند پرداخت کرایه اتوبوس و مترو شهر اصفهان می‌باشد.

۲- مرور ادبیات

در سال‌های اخیر و با توسعه استفاده از دستگاه‌های خودکار پرداخت کرایه در حمل‌ونقل همگانی داده‌های بسیاری ثبت و جمع‌آوری شده است. به‌منظور استفاده هرچه بهتر داده‌های ثبت‌شده در مدیریت و برنامه‌ریزی حمل‌ونقل، محققان مطالعات متعددی را در این حوزه انجام و روش‌هایی را برای تجزیه‌وتحلیل این اطلاعات توسعه داده‌اند. به‌طورکلی مطالعات انجام‌شده پیرامون تخمین ماتریس مبدا-مقصد عموماً شامل مراحل پاکسازی داده، شناسایی سفرهای دارای انتقال از گونه‌ای به گونه دیگر، تخمین موارد نامشخص (ایستگاه‌های سوار و پیاده شدن) و اعتبارسنجی روش‌های ارائه شده است[۱] . به‌طورکلی سامانه‌های خودکار پرداخت کرایه به دو نوع فقط ورود و ورود و خروج تقسیم می‌شوند. در دسته اول مسافران فقط هنگام ورود از کارت خود برای پرداخت استفاده کرده و تمام مسیر دارای کرایه ثابت است. در مقابل در گروه دوم با کارت زدن در هنگام سوارشدن و پیاده شدن میزان کرایه را بر اساس میزان مسافت پیموده شده پرداخت می‌کنند که در این صورت محل سوار و پیاده شدن مسافران در دسترس خواهد بود. در دسته اول با توجه به

[۱] Automatic Fare Collection (AFC)

اینکه مسافران در هنگام خروج از کارت خود استفاده نمی‌کنند، ایستگاه پیاده شدن نیز نامشخص بوده و قبل از تخمین مبدا-مقصد مسافران باید برآورد گردد.

- برآورد محل سوارشدن

بر اساس ماهیت و ویژگی‌های داده‌های ثبت‌شده اکثر مطالعات بر روی برآورد ایستگاه مقصد متمرکز بوده است. بااین‌حال تعیین ایستگاه سوارشدن نیز یکی از چالش‌های موردبررسی در برخی مقالات بوده است. ایتیکاف و همکاران [۱] روش‌های مختلف مورداستفاده برای برآورد ایستگاه سوارشدن را بررسی کرده که خلاصه آن در جدول ۱-۲ نشان داده شده است. بر این اساس در صورت در دسترس نبودن مکان سوارشدن، موقعیت آن با استفاده از منابع داده‌ای دیگر نظیر داده‌های AVL برآورد می‌شود.

جدول ۱-۲: روش‌های مورداستفاده با توجه به داده‌های در دسترس

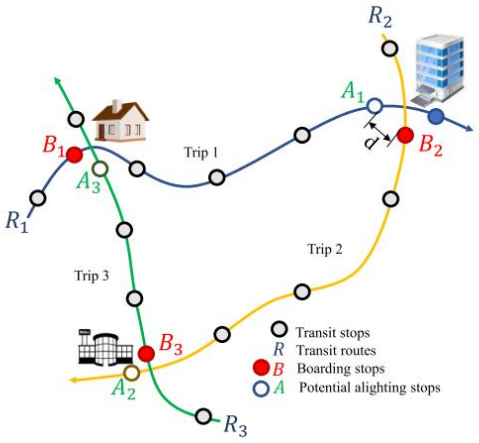
زمان سوارشدن	ایستگاه سوارشدن	توضیحات
√	√	هیچ محاسبه‌ای برای برآورد ایستگاه سوارشدن نیاز نیست.
√	×	موقعیت سوارشدن می‌تواند با ترکیب داده‌های کارت هوشمند و سایر مجموعه داده‌ها نظیر AVL، داده برنامه زمانی و ... در صورت در دسترس بودن انجام شود.
×	×	موقعیت سوارشدن می‌تواند با GTFS[۲]، APC، AFC بدست آید

- برآورد محل پیاده شدن

در بسیاری از مطالعات همچون[۲] ، [۳] و [٤] از روش زنجیره سفرها[۳] برای تعیین محل پیاده شدن در زمانی که تنها ایستگاه سوارشدن در دسترس می‌باشد، استفاده شده و محل پیاده شدن از طریق محل سوارشدن سفرهای متوالی استنباط می‌شود. در این روش تمامی سفرهای هر مسافر در طول روز به‌صورت زنجیره‌ای از سفرها در نظر گرفته‌شده و محل پیاده شدن از نظر مکان و زمان به مکان سوارشدن سفرهای متوالی محدود می‌شود. در شکل ۱-۲ مثالی از این روش ارائه شده است. در این شکل فردی ابتدا در ایستگاه B_1 سوار خط R_1 شده و پس از مدتی در خط R_2 و ایستگاه B_2 مشاهده شده است. با در نظر گرفتن نزدیک‌ترین ایستگاه خط R_1 به ایستگاه B_2، ایستگاه پیاده شدن (امکان‌پذیر) از خط R_1 برآورد می‌شود. این فرایند برای تمامی سفرها ادامه یافته و درنهایت و در سفر سوم با توجه به اینکه سفر دیگری بعدازآن وجود ندارد، با فرض اینکه فرد به محل اولیه شروع سفر بازگشته است (خانه یا محل کار یا...) ایستگاه پیاده شدن امکان‌پذیر در نزدیک‌ترین فاصله به ایستگاه سوارشدن اولین سفر برآورد می‌گردد. شایان‌ذکر است هر یک از مطالعات پیشین از معیارهایی مختلف و متناسب با ویژگی‌های داده، شرایط محیطی و نحوه عملکرد ناوگان برای تعیین آستانه‌های زمانی و مکانی استفاده کرده اند.

۳- روش انجام کار

روش انجام کار شامل چندین بخش است که در زیر به‌اختصار توضیح داده می‌شود. در بخش اول داده‌های ارائه‌شده پاکسازی شده و موردبررسی و تحلیل آماری قرار گرفتند. این بخش شامل بدست آورد زمان دقیق کارت زدن، حذف داده‌های تکراری و داده‌های پرت، ارزیابی اعتبار داده‌ها و غیره می‌باشد. درنهایت بعد از تهیه داده‌های نهایی، داده‌های اتوبوس و مترو با یکدیگر ترکیب و جدول ترکیبی تهیه گردید. در بخش دوم به‌منظور شناسایی و تفکیک خطوط مشابه، خوشه‌بندی اتوبوس‌ها بر اساس ویژگی‌هایی مانند تعداد مسافر (تعداد کارت‌های زده‌شده)، تعداد اتوبوس‌های هر خط و ... انجام شد. همچنین کارت‌های رانندگان و اپراتورها ایستگاه‌ها نیز شناسایی و از سایر کارت‌ها جدا شدند تا نتایج دقیق‌تری برای بررسی الگوی سفر مسافران بدست آید. با توجه به اینکه در داده‌های در اختیار قرارگرفته به‌جز مسافران مترو و بی آر تی ایستگاه سوارشدن سایر مسافران در خطوط معمولی مشخص نمی‌باشد، در بخش سوم از روش‌های مختلفی برای شناسایی و تخمین ایستگاه سوارشدن استفاده گردید. در بخش چهارم الگوریتمی جهت شناسایی ایستگاه‌های مبدأ و مقصد طراحی و مورداستفاده قرار گرفت و درنهایت با ناحیه بندی شبکه حمل‌ونقل همگانی، میزان تقاضای هر ایستگاه به ناحیه مربوطه اختصاص‌یافته و ماتریس مبدا-مقصد برآورد می‌شود. در ادامه به بیان توضیحاتی پیرامون هر یک از قسمت‌ها پرداخته شده است.



شکل ۱-۲: نمونه‌ای از برآورد ایستگاه با استفاده از زنجیره سفرها

۱-۳- پاکسازی‌داده‌ها (Data Cleaning)

منظور از پاکسازی داده‌ها، حذف و اصلاح داده‌های غلط از رکوردها و جداول درون بانک اطلاعاتی ارائه شده است. به عبارت دیگر هدف از پاکسازی داده، به دست آوردن اطلاعات معتبر، دقیق، کامل، سازگار و منحصربه‌فرد است. در ادامه به بررسی هرکدام از مراحل پاکسازی داده‌ها پرداخته شده است:

- **بدست آوردن زمان دقیق کارت زدن**

[۲] Google Transit Feed Specification

[۳] trip chaining method

مقادیر فیلد Time به‌صورت دقیق نمی‌توانست زمان کارت زدن را نمایش دهد، بنابراین با تغییر دادن فرمت فیلد Time به فرمت hh:mm:ss و قرار دادن آن در یک فیلد جداگانه به نام f_date زمان به‌صورت دقیق محاسبه گردید.

- **حذف داده‌های تکراری**

در هنگام جمع‌آوری داده‌ها ممکن است یکسری از رکوردها به‌صورت تصادفی دو یا چند بار ذخیره شوند و دارای ورودی‌های یکسان باشند، بنابراین داده‌ها و رکوردهای تکراری که باعث افزونگی در جداول و بروز خطا در نتایج خواهند شد می‌بایست حذف شوند. در مجموعه داده موردمطالعه نیز تعداد قابل‌توجهی رکورد تکراری در داده‌ها مشاهده گردید که در این مرحله شناسایی و حذف شد.

- **حذف داده‌های پرت**

در بررسی داده‌ها تعدادی از خطوط دارای داده‌های خارج از ساعات صحیح بودند. (با توجه به اینکه اکثر داده‌های اصلی اتوبوس و مترو در بین ساعات ۵ صبح تا ۱۰:۳۰ شب ذخیره شده بودند و همچنین با توجه به پرسش‌های تیم از واحدهای مختلف شهرداری و اتوبوسرانی و مترو و...). بنابراین داده‌های بین ساعت ۱۰:۳۰ شب تا ۵ صبح از جداول حذف شدند.

- **بررسی و تحلیل آماری**

در این قسمت تیم درصدد طراحی نمودارها و جداولی بود که توسط آن‌ها به یکسری الگوهای تکراری، روندها و ارزیابی اعتبار داده‌ها از منظر کسب‌وکار حمل‌ونقل در رکوردها دست یابد. این مرحله توسط نرم‌افزارهایی مانند DataGrid، Excel و Sql Management Studio انجام شد، که فایل‌ها و نتایج آن در پیوست ارائه شده است. (این مرحله روی داده‌های اولیه انجام‌شده است.)

- **ترکیب داده‌های اتوبوس با مترو و ساخت جدول ترکیبی**

در این مرحله داده‌های ۱۲ روز با یکدیگر ترکیب‌شده تا ضمن تجمیع و یکپارچه شدن داده‌ها از هم‌افزایی داده‌های مترو و اتوبوس برای برآورد ایستگاه‌های سوار و پیاده شدن استفاده شود. همچنین به‌منظور افزایش کارایی داده‌های اولیه و تأمین ویژگی‌های موردنیاز مراحل بعدی فیلدهای مختلفی به داده‌های اولیه اضافه گردید که شامل موارد زیر است:

- ✓ **فیلد isBRT** – با توجه به حجم بالای مسافران خطوط BRT نسبت به سایر خطوط اتوبوس، لازم است اطلاعات آن به‌صورت جداگانه موردبررسی قرار گیرد. بدین منظور یک فیلد به نام isBRT در جدول اتوبوس اضافه و خطوط BRT را از بقیه داده‌ها جدا گردید.لازم به ذکر است مطابق با توضیحات ارائه‌شده فیلد Plat که شامل پلاک اتوبوس‌ها است در خطوط عادی بیانگر پلاک خودرو و در خطوط BRT بیانگر ایستگاه خط می‌باشد. همچنین شماره خطوط BRT برابر خط ۶۹ و خط ۶ می‌باشد.
- ✓ **فیلد isMetro** – این فیلد برای شناسایی راحت‌تر رکوردهای مترو و اتوبوس تعریف و با مقدار صفر و یک نشان داده می‌شوند. درصورتی‌که‌صفر باشد بیانگر سفر با اتوبوس و در غیر این صورت سفر با مترو را نشان می‌دهد.
- ✓ **فیلد Station** - بر اساس نوع خطوط (BRT یا عادی و یا مترو) فیلدی به نام station ایجاد گردید که در اتوبوس‌های عادی برابر صفر، در BRT از عدد ۱ تا ۳۳ و در مترو از ۱۲۶ تا ۱۵۴ (طبق داده‌های تحویل داده‌شده) است. کدگذاری ایستگاه‌های مترو به شرح زیر است:

```
UPDATE Table SET station =
CASE
    WHEN station = 'Abshar' THEN 1
    WHEN station = 'Al Khujand' THEN 2
    WHEN station = 'Apadana' THEN 3
    WHEN station = 'Artesh' THEN 4
    WHEN station = 'Azadi' THEN 5
    WHEN station = 'BaghoshKhane' THEN 6
    WHEN station = 'Biston' THEN 7
    WHEN station = 'Emam sadeq' THEN 8
    WHEN station = 'Feriburg' THEN 9
    WHEN station = 'Golzar 7th' THEN 10
    WHEN station = 'GolzarShohada' THEN 11
    WHEN station = 'HashtBehesht' THEN 12
    WHEN station = 'HosseinAbad' THEN 13
    WHEN station = 'Keshavarzi' THEN 14
    WHEN station = 'Masjed Al Anbia' THEN 15
    WHEN station = 'Masjed Alghafur' THEN 16
    WHEN station = 'Mesami' THEN 17
    WHEN station = 'Molasadra' THEN 18
    WHEN station = 'Norbaran' THEN 19
    WHEN station = 'QasreGol' THEN 20
    WHEN station = 'Qazeh Jonob' THEN 21
    WHEN station = 'Qazeh shomal' THEN 22
    WHEN station = 'SapahSalar' THEN 23
    WHEN station = 'SheykhSadoq' THEN 24
    WHEN station = 'Simin' THEN 25
    WHEN station = 'South Baskoul' THEN 26
    WHEN station = 'South Qods' THEN 27
    WHEN station = 'Sq.AhmadAbad' THEN 28
    WHEN station = 'Sq.Bozorgmehr' THEN 29
    WHEN station = 'Taghato' THEN 30
    WHEN station = 'Tohid' THEN 31
    WHEN station = 'Zendan' THEN 32
    WHEN station = 'sayar' THEN 33
    ELSE 0
END;
```

- ✓ **فیلد commonUserInDay** – این فیلد مسافرانی که از هر دو گونه مترو و اتوبوس در یک روز استفاده کردند را با عدد یک نشان دهد.

- ✓ **فیلد isRepitedInDays** - بیانگر مسافرانی که در یک هفته بیشتر از یک‌بار از هردو گونه استفاده کردند (جهت تفکیک الگوهای تکرارشونده از سفرهای یک‌بار مشاهده‌شده)
- ✓ **فیلد countInDay** – تعداد استفاده از یک کارت آی‌دی منحصربه‌فرد در یک روز را نمایش می‌دهد.
- ✓ **فیلد uniqueAllDays** – آی‌دی‌هایی که در کل بازه موردمطالعه (۱۲ روز) فقط یک‌بار استفاده شده است را نمایش می‌دهد.
- ✓ **فیلد countLineCodeInADay** – تعداد لاین کدهای مختلف استفاده‌شده توسط هر کارت آی‌دی منحصربه‌فرد در هرروز را نمایش می‌دهد.
- ✓ **فیلد countLineCodeInAllDay –** تعداد لاین کدهای مختلف استفاده‌شده توسط هر کارت آی‌دی منحصربه‌فرد در ۱۲ روز را نمایش می‌دهد.
- ✓ **فیلد commonUserInDayMetroBRT** – کارت آی‌دی‌هایی که در یک روز هم مترو و هم BRT استفاده کرده‌اند را مشخص می‌کند.
- ✓ **فیلد isLocale –** این فیلد بیانگر خطوط اتوبوسی است که دارای طول کوتاهی هستند و به‌اصطلاح به‌صورت محلی مورداستفاده قرار می‌گیرند که شامل خط‌های ۵۳، ۵۰، ۵۶، ۵۸، ۳۶۵، ۴۶، ۱۵۰، ۳۸، ۱۴۸، ۲۹۱، ۴، ۹، ۱۵۷، ۶۳، ۳۳، ۱۹، ۱۷۶، ۱۸ است. لازم به ذکر است خطوط با طول کمتر از ۷ کیلومتر به‌عنوان خطوط محلی در نظر گرفته شده است.
- ✓ **فیلد newStation –** این فیلد ایستگاه سوارشدن مسافر را مشخص می‌کند و شامل شماره ایستگاهی است که مسافر در آن کارت زده و سوار شده است.
- ✓ **فیلد getOut –** این فیلد ایستگاه پیاده شدن را نمایش داده که با استفاده از توالی سفرها استنباط می‌شود.

۲-۳- خوشه‌بندی داده‌ها

با توجه به اینکه الگوهای زمانی و مکانی در اتوبوس‌های خلوت، معمولی و شلوغ متفاوت است، برای شناسایی و تفکیک الگوها خوشه‌بندی داده‌ها بر اساس موارد زیر انجام پذیرفت.

- **خوشه‌بندی بر اساس تعداد کارت‌های زده شده**

این خوشه‌بندی بر اساس تعداد مسافر در هر خط می‌باشد. خوشه‌ها به‌صورت زیر می‌باشد:

- ✓ ۰ – ۴۰۰۰ کارت
- ✓ ۴۰۰۰ – ۱۰۰۰۰ کارت
- ✓ ۱۰۰۰۰ – ۲۰۰۰۰ کارت
- ✓ ۲۰۰۰۰ تا ۳۰۰۰۰ کارت
- ✓ ۳۰۰۰۰ به بالا

- **خوشه‌بندی بر اساس تعداد اتوبوس**

این خوشه‌بندی بر اساس تعداد اتوبوس در هر خط است. تعداد اتوبوس از تعداد Plat های منحصربه‌فرد آن خط به دست آمده است.

- ✓ ۱- ۵
- ✓ ۵ – ۱۰
- ✓ ۱۰ – ۲۰
- ✓ ۲۰ - ۳۰
- ✓ ۳۰ به بالا

- **شناسایی و تفکیک کارت راننده – مسافر**

در بررسی داده‌ها مشاهده گردید تعدادی از کارت‌ها دارای تراکنش بیش‌ازحد معمول در هرروز هستند که احتمالاً این کارت‌ها مربوط به کارت‌های راننده و اپراتورهای مستقر در ایستگاه‌ها می‌باشد که در موقعیت‌های مختلف مورداستفاده قرار می‌گیرد (مثلاً در زمان‌هایی که مسافر کارت ندارد و وجه نقد استفاده می‌کند). با توجه به امکان بروز خطا در الگوهای سفر (در صورت استفاده از تراکنش‌های ثبت‌شده توسط این کارت‌ها)، لازم است این موارد شناسایی و از سایر کارت‌ها متمایز شود. بر این اساس فیلدی در کنار جدول اصلی قرار گرفت و کارت آی‌دی‌هایی که احتمال داده شد راننده هستند با مقدار ۱ در فیلد isDriver قرار گرفت. میزان درستی این ادعا توسط یکسری از فاکتورها بررسی گردید که در فایل‌های پیوست ارائه شده است. برای شناسایی این کارت‌ها از معیارهای متفاوتی استفاده گردید که اهم آن مطابق زیر است:

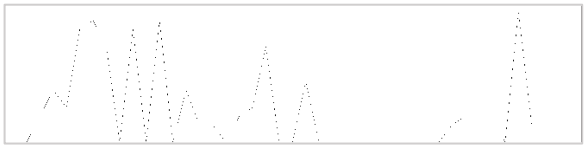
- ✓ آی‌دی کارت در یک روز و در یک Plat (پلاک) تکرار شده باشند.
- ✓ تعداد ثبت شده آن آی‌دی کارت از یک تعدادی بالاتر باشند. این تعداد توسط خوشه‌بندی خطوط که در بالا ذکر شد مشخص شد و در خطوط مختلف تفاوت دارد.
- ✓ اختلاف‌زمانی یک کارت آی‌دی بالاتر از حد مجاز باشد (مثلاً کارت آی‌دی انتخابی به‌صورت متوالی در یک ساعت زده شده باشد). این زمان نیز توسط خوشه‌بندی خطوط که در بالا ذکر شد مشخص گردید و در خطوط مختلف تفاوت دارد.

با کمک شروط بالا یکسری آی‌دی کارت به دست آمد که به‌احتمال‌زیاد کارت‌های رانندگان هستند. به‌منظور اطمینان از نتایج حاصل‌شده از یک روش دیگر برای تشخیص کارت راننده استفاده گردید. بر اساس این روش و با توجه به ارتباط میان کارت راننده و تعداد راننده هر اتوبوس و خط، انحراف معیار تعداد اتوبوس منحصربه‌فرد در هر خط و هرروز محاسبه گردید و میزان "میانگین ±Std۲ " به‌عنوان آستانه‌های تعداد تراکنش که مرز تراکنش‌های مسافران عادی و راننده را مشخص می‌کند حاصل و بر اساس آن کارت‌های رانندگان شناسایی و در ستونی دیگر برچسب زده شد. نتایج این روش نیز در فایل اکسل count_ID_for_driver ارائه شده است.

۳-۳- شناسایی و برآورد ایستگاه سوارشدن

با توجه به اینکه در داده‌های موجود صرفاً پلاک اتوبوسی که مسافر سوار آن شده و زمان سوارشدن (کارت زدن) وجود دارد، بنابراین برای تخمین ماتریس مبدا-مقصد می‌بایست ابتدا ایستگاه‌های سوار و پیاده شدن استنباط و برآورد شود. لذا در این مرحله ابتدا سعی بر آن شد تا ایستگاه‌های سوارشدن با استفاده از زمان حضور اتوبوس در ایستگاه و کارت زدن مسافران شناسایی شود. به‌عبارت‌دیگر در زمان رسیدن اتوبوس به ایستگاه، مسافران شروع به کارت زدن نموده و سوار می‌شوند و پس‌ازآن و در فاصله زمانی پیمایش تا ایستگاه بعدی

هیچ کارتی (یا کارت‌های متوالی) زده نمی‌شود. بر این اساس درصورتی‌که نمودار تعداد تراکنش‌های ثبت‌شده (کارت زده‌شده) در طول زمان رسم گردد، تراکنش‌های ثبت‌شده متوالی می‌تواند بیانگر ایستگاه سوارشدن و گپ‌های بین این تراکنش‌های متوالی بیانگر فاصله پیمایش بین دو ایستگاه باشد. با بررسی‌های انجام‌شده و ترسیم این نمودار در بازه‌های زمانی سه، پنج و پانزده دقیقه، مناسب‌ترین بازه زمانی برای نمایش و تشخیص ایستگاه با در نظرگیری فواصل بین ایستگاه‌ها و زمان پیاده و سوارشدن مسافران، بازه زمانی سه دقیقه بدست آمد. در شکل ۱-۳ نمونه‌ای از نمودار رسم شده برای یک اتوبوس در طول چهار ساعته که تقریباً برابر یک‌ونیم سیکل حرکت آن می‌باشد نشان داده شده است.



شکل ۱-۳: تعداد تراکنش در طول زمان (تجمعی ۳ دقیقه)

با توجه به اینکه زمان پیمایش فاصله بین دو ایستگاه در ساعات مختلف متفاوت بوده و به ترافیک معابر و ... وابسته است، ممکن است تجمیع تراکنش‌ها در بازه سه‌دقیقه‌ای در زمان‌های خلوتی شبکه نتایج واقع‌بینانه‌ای در بر نداشته باشد لذا برای رفع این مسئله روشی دیگر طراحی گردید که در آن فاصله زمانی بین تراکنش‌ها در هر ثانیه رسم شده و با محاسبه انحراف معیار فاصله بین کارت زدن‌ها و در نظرگیری ۲ یا ±Std۳ سعی در پیدا نمودن حداکثر زمان بین کارت زدن‌ها در هنگام سوارشدن گردید. پس از محاسبه حداکثر فاصله زمانی امکان‌پذیر بین دو تراکنش، درصورتی‌که زمان بین دو تراکنش از حداکثر تعیین‌شده بیشتر باشد بیانگر این است که اتوبوس به ایستگاه بعدی رفته است. این روش نیز علیرغم کارایی در برخی موارد کارا نبوده و نتوانست ایستگاه‌ها را به‌طور کامل مشخص نماید. اما در ادامه و برای رفع مشکلات این روش، نمودار لگاریتم فاصله‌های زمانی رسم گردید که نمونه از آن در شکل ۲-۳ نمایش داده شده است. با بررسی انجام‌شده در این نمودار احتمالاً هر پرش بیانگر یک ایستگاه خواهد بود.

شکل ۲-۳: نمودار فاصله زمانی بین تراکنش‌ها (لگاریتم)

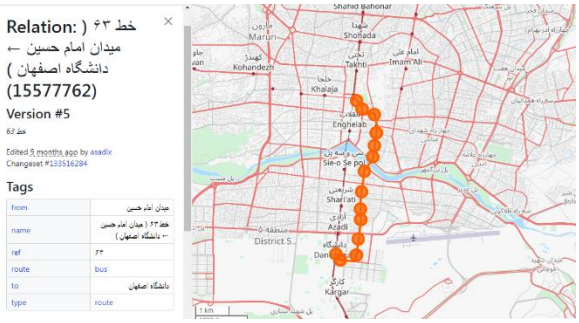
سپس به‌منظور بررسی بیشتر نمودارهای چند اتوبوس یک خط، در یک‌هشتم ابتدایی زمان کاری در یک روز رسم گردید که در پیوست ارائه شده است. ازآنجاکه خطوط در نزدیکی هم قرار گرفتند می‌توان نتیجه گرفت که این روش می‌تواند روشی مناسب برای پیدا کردن ایستگاه‌های سوارشدن باشد.

درعین‌حال با توجه به عدم قطعیت‌های موجود نظیر عدم وجود مسافر در یک ایستگاه، عدم توقف اتوبوس در ایستگاه، کارت زدن در فاصله پیمایش دو ایستگاه، خرابی اتوبوس و ...) و همچنین زمان‌بر بودن محاسبه روش ارائه‌شده، این روش مورداستفاده قرار نگرفت و در بخش بعدی روشی کارآمد و سریع‌تر بر اساس ویژگی‌های قطعی در دسترس ارائه و استفاده شده است.

۴-۳- شناسایی و برآورد هم‌زمان ایستگاه سوار و پیاده شدن

همان‌طور که ذکر شد با توجه به زمان‌بر بودن روش‌های قبلی و محدودیت زمانی چالش، نیاز به استفاده از یک روش سریع و با کمترین میزان عدم قطعیت وجود دارد. بر این اساس درروشی که طراحی گردید از ایستگاه‌های سوارشدن مترو و بی آر تی به‌عنوان داده‌های موجود و قطعی و از توالی سفرها و تلاقی خطوط مختلف با یکدیگر برای شناسایی ایستگاه‌های سوار و پیاده شدن استفاده گردید در ادامه به آن پرداخته شده است.

قبل از بدست آوردن ایستگاه‌های سوار و پیاده شدن، نیاز به در دسترس داشتن اطلاعات ایستگاه‌های مختلف است. بدین منظور برای به دست آوردن مختصات و ویژگی‌های خطوط و ایستگاه‌ها از داده‌های open Street Map استفاده شد و ایستگاه‌های تمامی خطوط به دست آمد.



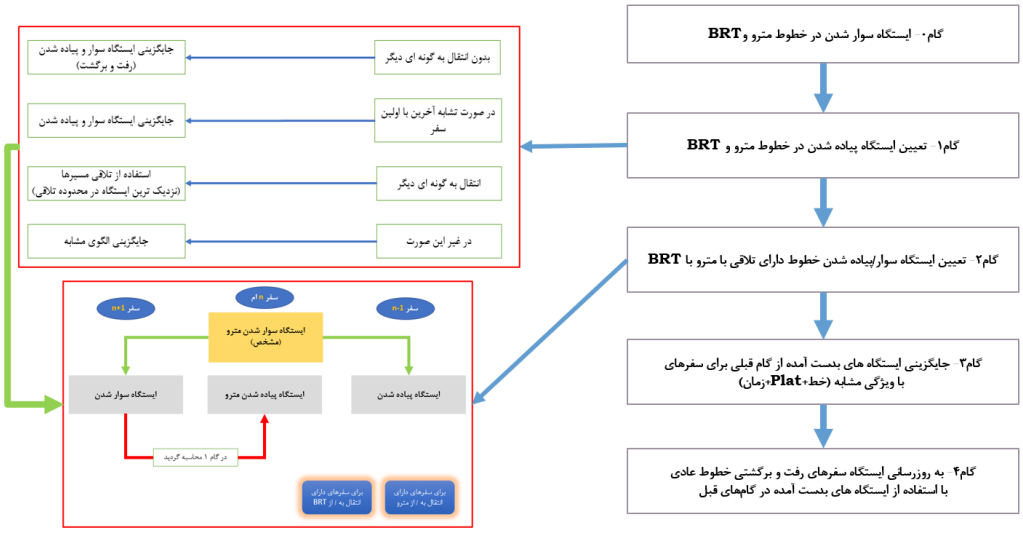
شکل ۳-۳: نمونه ایستگاه‌های بدست آمده از OSM

مطابق شکل ۴-۳ مراحل این روش در ۴ گام ارائه شده است. در گام نخست با در دست داشتن ایستگاه‌های سوارشدن سفرهای مترو و بی آر تی، به دنبال یافتن ایستگاه پیاده شدن آن سفرها هستیم. درصورتی‌که دو سفر متوالی یک آی دی بدون انتقال به‌گونه‌ای دیگر انجام‌شده باشد (بدون جابجایی از مترو به بی آر تی یا اتوبوس معمولی و بلعکس) ایستگاه سوارشدن سفر دوم به‌عنوان ایستگاه پیاده شدن سفر اول در نظر گرفته می‌شود. درصورتی‌که در سفر دوم انتقال به‌گونه‌ای دیگر صورت گرفته باشد از تلاقی خطوط سفر اول و دوم استفاده‌شده و نزدیک‌ترین ایستگاه خط اول به ایستگاه‌های خط دوم به‌عنوان ایستگاه پیاده شدن سفر اول و نزدیک‌ترین ایستگاه خط دوم به آن به‌عنوان ایستگاه سوارشدن سفر دوم لحاظ می‌شود. شایان‌ذکر است نزدیک‌ترین ایستگاه‌های خطوط دارای تلاقی در محیط GIS پیاده‌سازی و احصا گردید که نتایج آن در پیوست ارائه‌شده است.

همچنین درصورتی‌که صرفاً اولین و آخرین سفر در گونه مشابه انجام‌شده باشد از جابجایی ایستگاه‌های سوارشدن دو سفر برای ایستگاه پیاده شدن استفاده می‌گردد. در مواردی که هیچ‌یک از حالات ذکرشده اتفاق نیفتاده باشد می‌توان از جایگزینی الگوی افراد مشابه (افرادی که هم‌زمان در یک ایستگاه سوار یک اتوبوس شده‌اند) استفاده کرد.

در گام بعدی ایستگاه‌های سوار یا پیاده شدن مسافران به/از خطوط عادی اتوبوس با استفاده از تلاقی آن‌ها با خطوط مترو و بی آر تی و فاصله ایستگاه‌ها نسبت به یکدیگر تعیین می‌شود. در گام سوم برای سفرهایی که در دو مرحله قبلی هیچ ایستگاهی تعیین نشده است، الگوی سفر افراد مشابه جایگزین آن‌ها می‌شود. به‌عبارت‌دیگر با ایستگاه سوار و پیاده شدن افرادی که در یک‌زمان سوار اتوبوس یکسان شده‌اند (بدست آمده از مراحل قبلی) برای آن‌ها جایگزین

می‌شود. در گام آخر ضمن به‌روزرسانی ایستگاه‌های حاصل‌شده، مجدداً الگوریتم تکرار شده تا ایستگاه‌های نامشخص باقی‌مانده تعیین گردد.

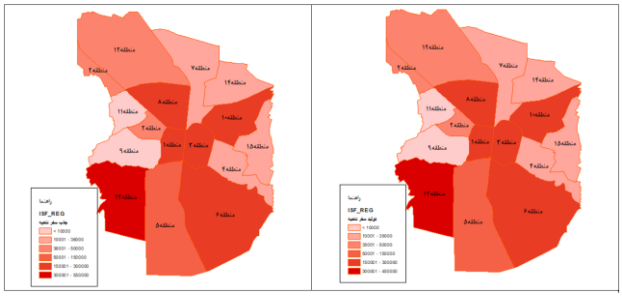


شکل ۴-۳: مراحل انجام پژوهش

٤- نتایج

پس از برآورد و تعیین ایستگاه‌های سوار و پیاده شدن، برای برآورد ماتریس مبدا-مقصد تعداد مسافر هر ایستگاه به ناحیه مربوطه تخصیص می‌یابد. بدین منظور از ۲ ناحیه بندی متفاوت استفاده‌شده که شامل نواحی پانزده‌گانه شهر اصفهان و دیگری شامل ناحیه بندی به تعداد ۵۰ ناحیه با متوسط مساحت ۹ کیلومترمربع می‌باشد. نحوه ناحیه بندی نیز در پیوست ارائه شده است.

پس از یک مرحله اجرای الگوریتم، از مجموع ۶۴۰۵۴۱۹ سفر ثبت‌شده در کل بازه موردمطالعه، برای بیش از ۹۹۱۳۰۹ سفر (۱۵درصد از کل سفرها) صرفاً ایستگاه پیاده شدن، ۱۶۸۰۴۴۹ سفر (۲۶ درصد) صرفاً ایستگاه سوارشدن و ۲۰۴۱۵۳۳ (۳۲ درصد) سفر نیز هم ایستگاه سوار و هم پیاده شدن حاصل گردید. با توجه به اینکه بیش از ۳۰ درصد از کارت‌های ثبت‌شده در هرروز فقط ۱ تراکنش داشته و دارای الگوهای تکرارشونده نیز نبوده است، الگوریتم فوق توانست مبدا-مقصد بیش از ۴۶ درصد از کل کارت‌هایی که بیش از ۱ بار مورداستفاده قرارگرفته است را شناسایی نماید. با توجه به نیاز به ماتریس دقیق جهت استفاده در ارزیابی چالش و همچنین در دست نبودن اطلاعاتی جهت ارزیابی کالیبره کردن مدل، الگوریتم مذکور صرفاً یک مرحله اجرا گردید تا ماتریس حاصل‌شده از دقت بالایی برخوردار باشد. بدیهی است در صورت تکرار الگوریتم، ایستگاه‌های سوار و پیاده شدن سفرهای بیشتری مشخص و تعیین خواهد شد. نتایج این الگوریتم به‌صورت ماتریس مبدا-مقصد و باقابلیت تفکیک بر اساس روز و گونه سفر به پیوست ارائه شده است. همچنین در شکل ۱-۴ میزان تولید و جذب سفر هر یک از نواحی نمایش داده شده است.



شکل ۱-۴: نتایج خروج (تولید و جذب سفر نواحی) برای مناطق پانزده‌گانه شهرداری

٥- جمع‌بندی و پیشنهادات

همان‌طور که اشاره شد با اجرای الگوریتم ارائه‌شده در یک مرحله مبدا-مقصد بیش از ۴۶ درصد از سفرها برآورد گردید که در صورت تکرار الگوریتم این میزان افزایش قابل‌توجهی داشته و عین حال نیازمند کالیبره نمودن زمانی و مکانی الگوریتم می‌باشد. همچنین در نظرگیری شروط زمانی (مانند حداکثر زمان انتقال) علاوه بر شروط مکانی (تلاقی خطوط) می‌تواند منجر به افزایش دقت نتایج گردد. در صورت در اختیار داشتن داده‌های تکمیلی دیگر (مانند AVL) نیز امکان اعتبارسنجی مدل ارائه‌شده وجود خواهد داشت. همچنین در ادامه این پژوهش نیز می‌توان با تمرکز بر روی الگوهای فردی و تعمیم آن به افراد و گروه‌های مشابه، دقت نتایج را بهبود بخشید.

منابع

[۱] A. B. E. C. Etikaf Hussain, "Transit OD matrix estimation using smartcard data: Recent developments and future research challenges," *Transportation Research Part C* , ۲۰۲۱.

[۲] A. A. M. M. M. H. M. M. Behrang Assemi, "Improving alighting stop inference accuracy in the trip chaining method using neural networks," *Public Transport*, ۲۰۲۰.

[۳] A. K. Q. H. Pramesh Kumar, "A robust method for estimating transit passenger trajectories using automated data," *Transportation Research Part C*, ۲۰۱۸.

[٤] J. A. N. H. M. W. Catherine Seaborn, "Analyzing Multimodal Public Transport Journeys in London with Smart Card Fare Payment Data," *Journal of the Transportation Research Board*, ۲۰۰۹.