

برآورد تعداد مسافر ایستگاه‌های خطوط اتوبوس شهر تهران: نمونه موردی خط پل کریمخان - میدان جمهوری

سید احسان سید ابریشمی^{۱*}، علی محدث دیلمی^۲، وجیهه امینی^۳، مریم ایران‌منش^۴

- ۱- سید احسان سید ابریشمی، دکتری برنامه‌ریزی حمل‌ونقل، استادیار دانشکده عمران و محیط زیست، دانشگاه تربیت مدرس، تهران
- ۲- علی محدث دیلمی، دکتری برق، گرایش مهندسی پزشکی، دانشکده برق و مهندسی کامپیوتر، دانشگاه تربیت مدرس، تهران
- ۳- وجیهه امینی، کارشناسی ارشد برنامه‌ریزی حمل‌ونقل، دانشکده عمران و محیط زیست، دانشگاه تربیت مدرس، تهران
- ۴- مریم ایران‌منش، کارشناسی ارشد برنامه‌ریزی حمل‌ونقل، دانشکده عمران و محیط زیست، دانشگاه تربیت مدرس، تهران

seyedabrishami@modares.ac.ir

تاریخ پذیرش ۹۷/۰۳/۲۳

تاریخ دریافت ۹۶/۰۵/۲۰

چکیده

با توجه به مشکلات ناشی از افزایش روزافزون جمعیت در شهرها و متوسط مالکیت وسیله نقلیه، تغییر در سیاست‌های حمل‌ونقل شهری و تلاش برای توسعه حمل‌ونقل عمومی، به ویژه اتوبوس، یکی از مهم‌ترین اقدامات در حوزه حمل‌ونقل شهری است. برنامه‌ریزی برای استفاده بهینه از زیرساخت‌های اتوبوسرانی نیازمند وجود اطلاعات در زمینه زیرساخت و تقاضای سفر ایستگاه‌های اتوبوس است. بر این اساس، انجام مطالعاتی برای برآورد تعداد مسافر ایستگاه‌های اتوبوس به منظور برنامه‌ریزی عملیاتی خطوط اتوبوسرانی در شهر تهران ضروری است. در این مطالعه با استفاده از دو پایگاه داده: داده‌های ثبت‌شده تراکنش‌های کارت‌بلیت‌ها و داده‌های ثبت‌شده توسط موقعیت‌یاب خودکار موجود در اتوبوس‌ها و ساخت مدل مناسب، تعداد مسافر ایستگاه‌های اتوبوس به منظور برنامه‌ریزی عملیاتی خطوط اتوبوسرانی در شهر تهران ضروری است. در این مطالعه با استفاده از دو پایگاه داده: داده‌های ثبت‌شده تراکنش‌های کارت‌بلیت‌ها و داده‌های ثبت‌شده توسط موقعیت‌یاب خودکار موجود در اتوبوس‌ها و ساخت مدل مناسب، تعداد مسافر ایستگاه‌های اتوبوس برای آینده کوتاه‌مدت پیش‌بینی می‌شود. ابتدا اطلاعات یاد شده مربوط به هر خط اتوبوس مرتب شده و سپس، برای تعیین ماتریس مبدأ-مقصد مسافران، دو پایگاه داده با استفاده از کدنویسی در نرم‌افزار متلب تطبیق داده می‌شود. بدین منظور، با استفاده از شماره پلاک اتوبوس و زمان تراکنش کارت‌بلیت، ایستگاه محل پیاده شدن مسافر تعیین شده و برای تخمین ایستگاه محل سوار شدن مسافران اتوبوس، تراکنش کارت‌بلیت‌های هوشمند ثبت شده به چندین دسته شامل داده‌های ثبت شده بدون شماره سریال کارت‌بلیت، داده‌های ثبت شده کارت‌بلیت‌های هوشمند با یکبار استفاده در طول شبانه‌روز، داده‌های ثبت شده کارت‌بلیت‌های هوشمند با دوبار استفاده در طول شبانه‌روز و داده‌های ثبت شده کارت‌بلیت‌های هوشمند با بیش از دو بار استفاده در طول شبانه‌روز تقسیم‌بندی شده و برای هر یک از آنها سیاستی جداگانه اجرا شده است. پس از تهیه ماتریس مبدأ-مقصد مسافر، این ماتریس به عنوان پایگاه داده مطالعه مورد نظر قرار گرفته و بر اساس آن مدل ساریما و شبکه عصبی پرداخت می‌شوند. نتایج مدل نشان می‌دهد، مدل پروسپترون چند لایه از لحاظ شاخص‌های خطا در پیش‌بینی تعداد مسافر ایستگاه‌های اتوبوس هر خط از مدل ساریما برتر بوده و روش مناسب‌تری برای برآورد تعداد مسافر ایستگاه اتوبوس است.

واژگان کلیدی: تقاضای مسافر ایستگاه اتوبوس، پیش‌بینی کوتاه‌مدت، سری زمانی، ساریما، شبکه عصبی، پرسپترون چند لایه

۱- مقدمه

برای انجام سفرهای کاری و دیگر سفرهای روزمره است. پیاده‌سازی و به کارگیری راهکارهای افزایش کارایی سیستم اتوبوسرانی موجود در دنیا شامل، زمان‌بندی ناحیه‌ای، ارسال

سیستم اتوبوسرانی یکی از بخش‌های مهم حمل‌ونقل بوده و عهده‌دار جابه‌جایی تعداد زیادی از افراد و اقشار جامعه

پیش‌بینی مدل شبکه عصبی نسبت به مدل اریما بیانگر، دقت بهتر مدل شبکه عصبی است. ژو و همکاران با استفاده از داده‌های حاصل از تراکنش کارت‌های هوشمند از آگوست تا نوامبر ۲۰۱۳، به پیش‌بینی کوتاه‌مدت تقاضای سفر با اتوبوس در یکی از خطوط شهر سنژن^۱ چین، پرداختند [3]. در این مطالعه، سه مدل پیش‌بینی هفتگی، روزانه و پانزده دقیقه‌ای برای بررسی ویژگی‌های متفاوت سری‌های زمانی پرداخت شده و به منظور ساخت مدل مناسب برای هر نوع داده، آزمایش ایستا^۲ انجام شده است. بدین ترتیب، با توجه به تناوب ثابت سری زمانی هفتگی مدل ارما^۳، تغییرات دوره‌ای سری زمانی روزانه مدل ساریما^۴ و نوسان گسترده تقاضای سفر در بازه‌های زمانی پانزده دقیقه طی یک روز، مدل اریما-گرچ^۵ (به تفکیک روزهای کاری و روزهای پایانی هفته) انتخاب شد. سپس با توجه به متفاوت بودن مشخصات سری زمانی هفتگی، روزانه و پانزده دقیقه‌ای و عدم امکان ساخت مدل واحد که پیش‌بینی دقیقی از ویژگی‌های موقعیت‌های مختلف داشته باشد، الگوریتم مدل چندگانه تعاملی^۶ بکار گرفته شد و نتایج آن، بیانگر عملکرد بهتر نسبت به مدل‌های جداگانه است. در مطالعه‌ای دیگر، بومن و همکاران به پیش‌بینی تقاضای سفر با اتوبوس در رتردام هلند پرداختند [4]. اطلاعات مورد استفاده در این مطالعه شامل داده‌های حاصل از تراکنش ۴ ماه کارت‌های هوشمند بوده و مدل‌سازی تقاضا با استفاده از سه روش سفر کوتاه-مبنا^۷، سفر-مبنا^۸ و الگوی سفر-مبنا^۹ است. بدین ترتیب، سفرهایی که شامل تنها یک بار استفاده از اتوبوس بوده در گروه سفر کوتاه-مبنا، سفرهایی که شامل چندین استفاده متوالی از اتوبوس هستند به طوری که مقصد یکی، مبدا دیگری می‌شود به صورت سفر-

ناوگان خالی، سفرهای گردش کوتاه و پرش ایستگاهی؛ نیازمند اطلاعات دقیقی از تقاضای سفر ایستگاه‌ها در بازه‌های زمانی کوتاه‌مدت است. بنابراین پیش‌بینی دقیق تقاضای سفر با اتوبوس برای هر ایستگاه، اهمیت زیادی در برنامه‌ریزی کوتاه‌مدت و زمان‌بندی اتوبوسرانی داشته و موجب بهبود عملکرد و اعتماد به خدمات این شیوه از حمل‌ونقل همگانی می‌شود.

اهمیت پرداختن به این موضوع در تهران، کلان‌شهری با بیش از ۸ میلیون سکنه و ۱۸ میلیون سفر روزانه که شهروندانش روزانه میلیون‌ها ساعت از وقتشان در ترافیک تلف شده و علاوه بر هزینه‌های سنگین مادی به مردم و دولت، عواقب و پیامدهای روحی و روانی غیرقابل محاسبه‌ای داشته، دوچندان است. گزیده آمار و اطلاعات حمل‌ونقل شهری در تهران در سال ۱۳۹۳ نشان می‌دهد، تقریباً ۲۱/۸ درصد از سفرهای انجام شده در این شهر با استفاده از تاکسی و ون، ۱۶/۴ درصد مترو، ۲۰ درصد توسط اتوبوس و مینی‌بوس؛ و ۴۱/۸ درصد خودروی شخصی و سایر وسایل انجام می‌گیرد [1]. بنابراین یکی از راهکارها برای رهایی از معضل ترافیک و مشکلات ناشی از آن، توسعه و بهبود عملکرد سیستم اتوبوسرانی است. این امر نیازمند برنامه‌ریزی و دسترسی به اطلاعات دقیق بوده و مهم‌ترین اطلاعات مورد نیاز در این زمینه برآورد دقیق تقاضای سفر با اتوبوس برای ایستگاه‌های مختلف است.

مرور مطالعات پیشین نشان می‌دهد، تاکنون مطالعات اندکی در زمینه پیش‌بینی کوتاه‌مدت تقاضای سفر اتوبوس در جهان انجام شده، برای نمونه، سلبی و بایراکتار به منظور پیش‌بینی کوتاه‌مدت تقاضای مسافران حمل‌ونقل همگانی در استانبول ترکیه از مدل شبکه عصبی استفاده نمودند [2]. در این مطالعات، داده‌های حاصل از تراکنش کارت‌های هوشمند در بازه زمانی جولای ۲۰۰۷ الی اکتبر ۲۰۰۸ برای آموزش شبکه، نوامبر ۲۰۰۸ برای ارزیابی کلی توانایی شبکه در طول آموزش و دسامبر ۲۰۰۸ به منظور آزمایش توانایی شبکه مورد استفاده قرار گرفت. داده‌های مورد استفاده در این مطالعه، به بازه‌های زمانی پانزده دقیقه‌ای تقسیم شده و مقایسه نتایج

1 Shenzhen

2 Stationary test

3 Autoregressive Moving Average (ARMA)

4 Seasonal Autoregressive Integrated Moving Average (SARIMA)

5 ARIMA-GARCH

6 Interactive Multiple Model algorithm

7 Trip-based

8 Tour-based

9 Pattern-based

یانتای^۸ چین است. به منظور پیش‌بینی تعداد مسافر اتوبوس در ایستگاه و زمان مشخص، ابتدا نمودار سری زمانی هر ایستگاه اتوبوس که شامل تعداد تقاضای مسافر است، ساخته شده، سپس سه مدل تغییرات زمانی پواسون^۹، مدل وزن‌دهی تغییرات زمانی پواسون^{۱۰} و مدل اریما ارائه شد. در نهایت، یک چارچوب کلی وزنی^{۱۱} برای پیش‌بینی تعداد مسافر با یک مدل واحد با استفاده از اطلاعات سری زمانی، ارائه شد. نتایج مطالعه بیانگر، عملکرد بهتر مدل کلی وزنی نسبت به سه مدل دیگر است.

همچنین فرانکنا در مطالعه‌ای به تخمین تجربی تابع تقاضای اتوبوس شهری در کانادا پرداخته؛ در این مطالعه فرض شده که سرانه تعداد تقاضای اتوبوس در یک ناحیه شهری به قیمت و هزینه زمان سفر با اتوبوس و خودروی شخصی، میانگین درآمد و ویژگی‌های اقتصادی-اجتماعی و مشخصات جغرافیای نواحی شهری مرتبط است. اطلاعات مورد نیاز مطالعه از اطلاعات سالیانه سیستم اتوبوسرانی شهری کانادا، کتاب‌ها و سازمان‌های مختلف تهیه شده و تابع تقاضا با استفاده از روش حداقل مجموع مربعات دو مرحله‌ای^{۱۲} تخمین زده شده است. نتایج مدلسازی نشان داد، با افزایش کرایه، تقاضا اتوبوس کاهش یافته، افزایش هزینه‌های مالکیت خودرو تأثیری بر تقاضای اتوبوس نداشته اما با افزایش قیمت بنزین تقاضای سفر با اتوبوس بیشتر می‌شود [8].

با توجه به آنکه در حال حاضر برنامه‌ریزی عملیاتی خطوط اتوبوسرانی شهر تهران همچون تعیین تناوب، زمان‌بندی اتوبوس و کارکنان، به صورت تجربی توسط مدیر خط انجام شده و برنامه‌ریزی بر اساس پیش‌بینی تقاضای آتی، وجود ندارد. این مطالعه به پیش‌بینی تعداد مسافر ایستگاه اتوبوس خطوط مختلف شهر تهران در افق کوتاه‌مدت (روزانه) با استفاده از اطلاعات AFC و AVL موجود در سیستم اتوبوسرانی پرداخته است.

مبنا و سفرهای که عضو این دو دسته نیستند به صورت الگوی سفر-مبنا مورد مدلسازی قرار گرفتند.

همچنین، سیپریچ و همکاران با استفاده از اطلاعات شمارش کرایه کامل برای هر مسافر که توسط متصدیان در شهر زیلینا^۱ اسلواکی^۲ جمع‌آوری شده و نرم‌افزار سس^۳، به پیش‌بینی تقاضای مسافران اتوبوس حومه شهری پرداختند [5]. مدل‌های مورد استفاده شامل باکس-جنکینس، یکنواخت‌نمایی، رگرسیون خطی چندگانه، سارما و مدل اریمای^۴ تغییر شکل یافته لگاریتمی بوده و نتایج نشان داد، عملکرد مدل اریمای تغییر شکل یافته بهتر از سایر مدل‌ها است. ما و همکاران، با استفاده از داده‌های تراکنش ثبت شده کارت‌های هوشمند یکی از خطوط شهر جینان^۵ چین که در طول یک سال جمع‌آوری شده به پیش‌بینی تقاضای کوتاه‌مدت مسافران اتوبوس پرداختند [6]. اطلاعات سری زمانی به صورت الگوی هفتگی، روزانه و ساعتی در نظر گرفته شده و با توجه به تست ایستا برای اطلاعات هفتگی مدل ار^۶، اطلاعات روزانه مدل سارما و اطلاعات ساعتی، مدل اریما انتخاب و برای روزهای کاری و پایانی هفته به صورت جداگانه پرداخت شد. با توجه به ویژگی‌های خاص الگوی تقاضا در بازه‌های زمانی مختلف، از مدل شبکه عصبی مصنوعی با الگوی هیبریدی^۷ و مدل چندگانه تعاملی با الگوی هیبریدی برای ساخت مدل واحد استفاده شد و مشاهده شد، مدل چندگانه تعاملی نسبت به سایر مدل‌ها، پیش‌بینی بهتری ارائه می‌کند. از طرفی ژوا^۸ و همکاران، سیستم پیش‌بینی تقاضای سفر اتوبوس با استفاده از اطلاعات سیستم واقعی اتوبوس را ارائه نمودند [7]. داده‌های مورد استفاده در این مطالعه شامل، اطلاعات ارسالی ۱۳۲۶ اتوبوس مجهز به دستگاه جی‌پی‌اس در ۴۱۶ ایستگاه اتوبوس بوده و مدت زمان برداشت اطلاعات یک دوره ۲۲ هفته‌ای بدون وقفه، در شهر

- 1 Zilina
- 2 Slovak
- 3 SAS (Statistical Analysis System)
- 4 Autoregressive Integrated Moving Average (ARIMA)
- 5 Jinan
- 6 Autoregressive (AR)
- 7 Artificial neural network hybrid pattern (ANNHP)

8 Yantai

9 Time Varying Poisson Model

10 Weighted Time Varying Poisson Model

11 weighted ensemble

12 Two stage least squares (TSLs)

بر این اساس وقتی پلاک اتوبوس (کد دستگاه کارت‌خوان اطلاعات AFC و شماره اتوبوس اطلاعات AVL) در هر دو دسته اطلاعات یکسان بوده و تراکنش ثبت شده از اطلاعات AFC کمترین اختلاف زمانی را با تراکنش ثبت شده از داده‌های AVL داشته باشد، فرض می‌شود که مسافر (صاحب کارت‌بلیت متناظر با آن تراکنش) در ایستگاهی پیاده شده، که تراکنش پیدا شده با کمینه اختلاف زمانی در آن ایستگاه بوده است. در شکل (۱) فلوچارت نحوه تطابق اطلاعات AFC و AVL نشان داده شده است. از آنجایی که زمان پرداخت کرایه توسط مسافران اتوبوس خطوط فیدر شهر تهران هنگام پیاده شدن از اتوبوس است، رکورد پیدا شده از مجموع اطلاعات AVL نشان‌دهنده ایستگاهی است که در آن مسافر، در مسیر معین از اتوبوس مشخص، پیاده شده است. توجه به این موضوع لازم است، در صورتی که اطلاعات ثبت شده توسط تراکنش کارت‌بلیت (سیستم AFC) در سیستم AVL وجود نداشته باشد، این دسته از اطلاعات حذف خواهند شد.

۲-۱- تخمین محل سوار شدن مسافران

تطابق اطلاعات AFC و AVL با استفاده از چند فرض، تنها نشان داد که هر مسافر در چه زمانی از روز، در کدام خط، از چه اتوبوسی و در کدام ایستگاه، از اتوبوس پیاده شده است. بنابراین، باید روشی برای تخمین محل سوار شدن مسافران اتوبوس ارائه شود. برای این منظور، با استفاده از چندین فرض که از مطالعات تطبیقی دیگر کشورها الهام گرفته شده، تراکنش کارت‌بلیت‌های هوشمند ثبت شده به چندین دسته تقسیم شده و برای هر یک از آنها سیاستی جداگانه اجرا شده است.

۲-۱-۱- داده‌های ثبت شده بدون شماره سریال کارت‌بلیت

داده‌های ثبت شده بدون شماره سریال کارت‌بلیت مربوط به آن دسته از مسافرانی است که به جای استفاده از کارت-بلیت هوشمند از وجه نقد برای پرداخت کرایه استفاده نموده و راننده اتوبوس از کارت خود به جای آن مسافر برای پرداخت کرایه استفاده می‌نماید.

ساختار مقاله بدین شکل است که پس از بیان مقدمه و مرور ادبیات در این بخش، در بخش دوم و سوم به ترتیب روش‌شناسی و تحلیل اطلاعات ارائه شده است. در بخش چهارم به بیان نتایج پرداخت مدل و بحث در این موضوع اشاره شده، در بخش پایانی (پنجم) نیز، نتیجه‌گیری و پیشنهادات بیان شده است.

۲- روش‌شناسی

اطلاعات موجود و ثبت شده توسط سیستم اتوبوسرانی شهر تهران شامل دو دسته؛ سیستم جمع‌آوری خودکار کرایه^۱ و سیستم موقعیت‌یاب خودکار وسیله نقلیه^۲ است. اطلاعات AFC (جمع‌آوری خودکار کرایه)، داده‌های ثبت شده حاصل از تراکنش‌های کارت‌بلیت‌های هوشمند برای پرداخت کرایه بوده که در اتوبوس فیدر^۳ و سامانه اتوبوس تندرو^۴ راه‌اندازی شده است. اطلاعات AVL نیز (داده‌های ثبت شده توسط موقعیت‌یاب خودکار) موقعیت جغرافیایی ناوگان اتوبوسرانی را با دقت قابل قبولی به مرکز اتوبوسرانی ارسال می‌کند. در جدول ۱ و جدول ۲ به ترتیب نمونه‌ای از اطلاعات AFC و AVL ارائه شده است.

به منظور برآورد تعداد مسافر ایستگاه‌های اتوبوس باید اطلاعات AFC و AVL در خطوط اتوبوسرانی فیدر شهر تهران، تطابق داده شود که در این بخش امکان و حالت‌های تطابق اطلاعات بررسی می‌شود. با توجه به اینکه در حال حاضر زمان ثبت شده در تراکنش‌های این دو سیستم هوشمند، با یکدیگر تطابق کامل نداشته و اختلاف زمانی میان داده‌های این دو سیستم مقدار ثابتی نیست، این موضوع از اولین چالش‌های تطابق اطلاعات AFC و AVL است. لازم به ذکر است، این اختلاف زمانی در بیشتر مواقع از سه دقیقه تجاوز نمی‌کند، ولی با توجه به متفاوت بودن اختلاف زمانی در هر ایستگاه و در هر ساعت روز، ناگزیر باید از تقریب‌هایی استفاده نمود که تا حد امکان معقول و اجرایی باشد.

- 1 Automated Fare Collection (AFC)
- 2 Automatic Vehicle Location (AVL)
- 3 Feeder bus
- 4 Bus rapid transit (BRT)

جدول ۱. نمونه‌ای از اطلاعات AFC دریافتی

Condition	Type smart cart	Serial number smart cart	Bus number plaque	Time use	Date use
valid	109	901668372	2821968	7:43	06/08/1394

Table 1- An example of AFC information

جدول ۲. نمونه‌ای از اطلاعات AVL دریافتی

Travel time (s)	Exit time from station	Arriveal time to station	Exit station code	Entrance station code	GPS serial	Bus number plaque	Date
220	12:45:12	12:48:52	7	8	1714547	420964711	09/04/1394

Table 2 An example of AVL information

یک خط از کارت استفاده کرده‌اند. در این داده‌ها، شماره سریال مسافری که از اتوبوس پیاده شده، مشخص است؛ اما چون تنها یکبار این شماره سریال ثبت شده است، مانند بخش قبل نمی‌توان با دقت مناسبی محل سوار شدن مسافر مربوطه را شناسایی کرد و فقط محل ایستگاه پیاده شدن مسافر از اتوبوس مشخص است. در این بخش نیز باید، این دسته از داده‌ها را در انتهای برآوردها، میان ایستگاه‌های قبل از محل سوار شدن مسافر به صورت نسبی با توجه به سفرهای غالب مسافران بدست آمده، توزیع نمود. در شکل ۲ (۲)، فلوچارت چگونگی توزیع سفر کارت‌بلیت‌هایی که تنها یکبار در طول شبانه روز مورد استفاده قرار گرفته، برای برآورد تعداد مسافرانی که در هر ایستگاه از اتوبوس سوار و پیاده می‌شوند، نشان داده شده است.

از آنجایی که این مسافران، شماره سریالی ندارند، ردیابی سفر آنان و تخمین مناسب محل سوار شدن آن‌ها به اتوبوس دور از دسترس است؛ ولی از آنجایی که محل پیاده شدن مسافر از اتوبوس مشخص است و قطعا مسافر یاد شده قبل از پیاده شدن از اتوبوس، در همان جهت مسیر به اتوبوس سوار شده است، می‌توان این دسته از داده‌ها را در انتهای این برآوردها، میان ایستگاه‌های قبل از محل سوار شدن مسافر به صورت نسبی با توجه به سفرهای غالب مسافران بدست آمده، توزیع نمود.

۲-۱-۲- داده‌های ثبت شده کارت‌بلیت‌های هوشمند با یکبار استفاده در طول شبانه‌روز داده‌های ثبت شده توسط سیستم AFC، مربوط به آن دسته از مسافرانی است که در طول شبانه‌روز تنها یکبار در

شکل ۱. چگونگی تطابق اطلاعات AVL و AFC

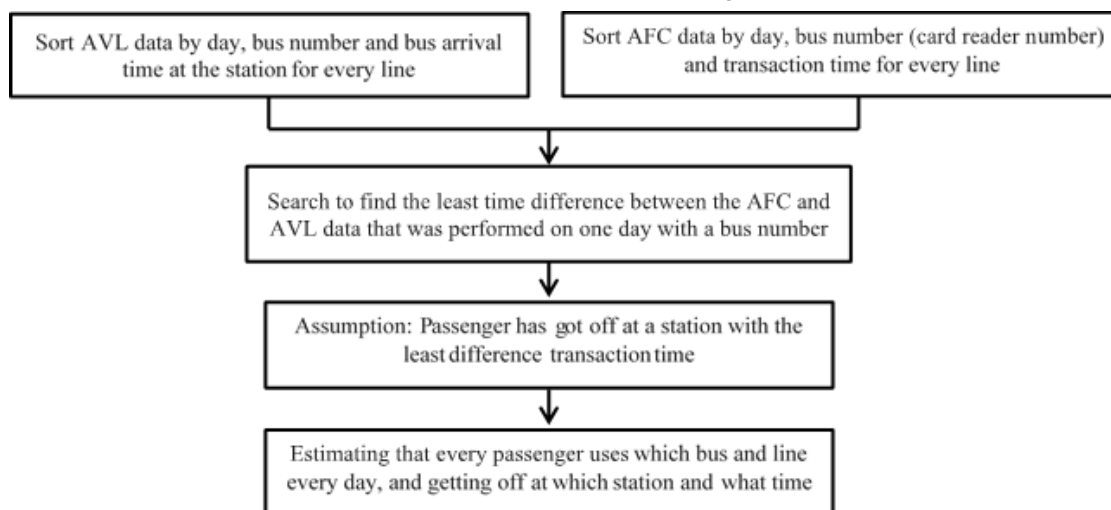


Fig 1. Method of match AFC and AVL information

شکل ۲. چگونگی توزیع سفر کارت‌بلیت‌ها با یکبار استفاده در طول شبانه روز

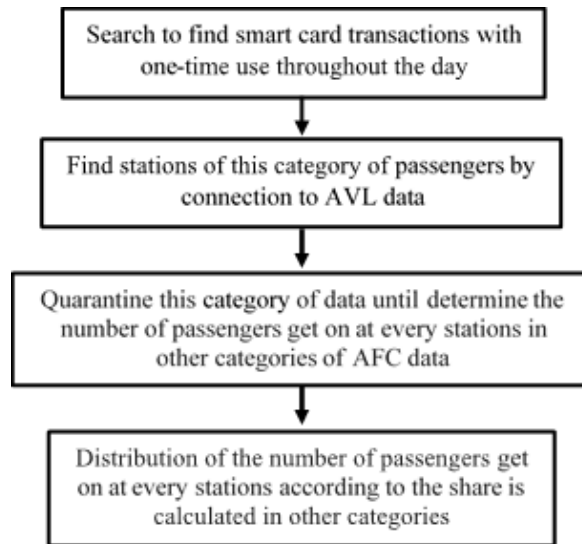


Fig. 2. Distribution method of travel smart cards with one-time use throughout the day

شکل ۳. چگونگی توزیع سفر کارت‌بلیت‌ها با دو بار استفاده در طول شبانه روز

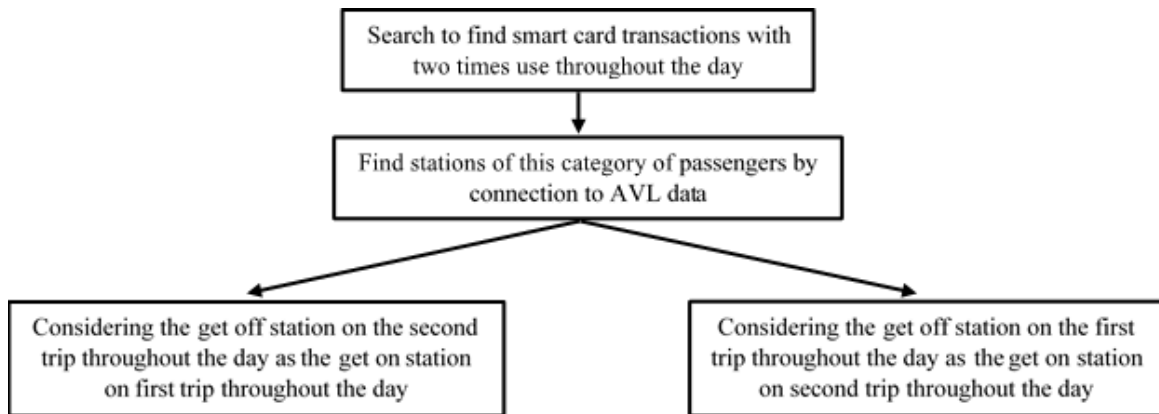


Fig 3. Distribution method of travel smart cards with two times use throughout the day

شکل ۴. چگونگی توزیع سفر کارت‌بلیت‌ها با بیش از دو بار استفاده در طول شبانه روز

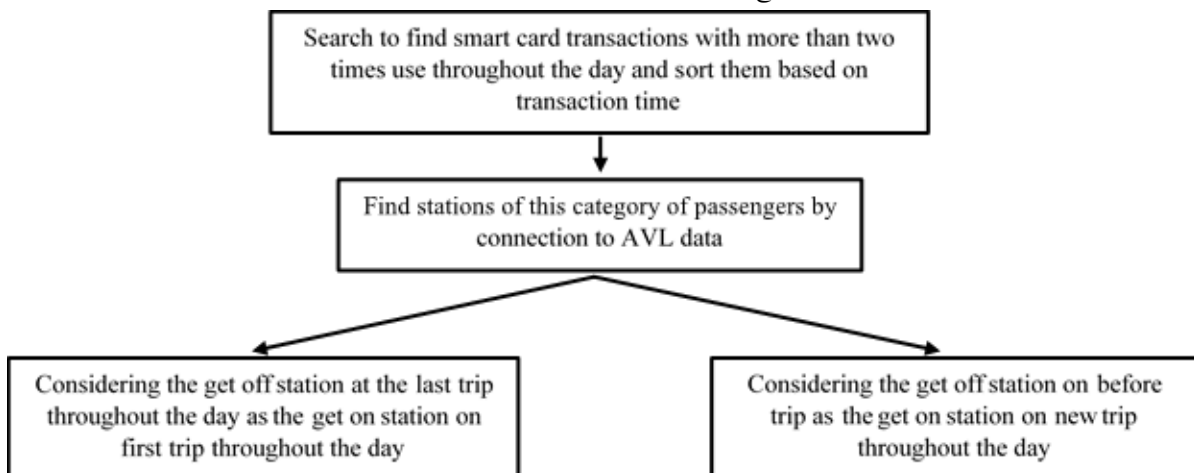


Fig. 4. Distribution method of travel smart cards with more than two times use throughout the day

و از ایستگاه پیاده شده باشد، سفر خود را از ایستگاهی آغاز می‌کند که در سفر قبل خود، از آن ایستگاه پیاده شده بوده ولی برای پیش‌بینی مبدأ سفر اول از توزیع بدست آمده برای ایستگاه‌های سوار شدن مسافران استفاده خواهد شد.

۲-۲- مدلسازی

پس از تطابق اطلاعات دو سیستم AFC و AVL، به منظور برآورد تعداد مسافر ایستگاه‌های خطوط اتوبوس از مدل‌های سری زمانی استفاده شده است. سری زمانی مجموعه منظمی از اطلاعات نقطه‌ای است که نوعاً روی زمان‌های متوالی سنجیده می‌شود. از نظر ریاضی، سری زمانی به صورت یک مجموعه برداری $x_t, t = 0, 1, 2$ تعریف شده که t معرف زمان منقضی شده و متغیر x_t به عنوان یک متغیر تصادفی^۱ عمل می‌نماید. هدف اصلی این مدل‌ها، گردآوری دقیق و مطالعه قوی مشاهدات گذشته یک سری زمانی به منظور توسعه یک مدل مناسب برای توصیف ساختار ذاتی یک سری است. از این مدل‌ها برای تولید مقادیر آینده یک سری زمانی به منظور پیش‌بینی استفاده می‌شود. بطور کلی، مدل‌های مورد استفاده برای سری‌های زمانی به دو دسته مدل‌های تصادفی^۲ و شبکه عصبی مصنوعی^۳ تقسیم‌بندی می‌شود [9].

۲-۲-۱- مدل‌های تصادفی

یک مدل سری زمانی با توجه به نوع اطلاعات کنونی که می‌تواند تابعی خطی و غیرخطی از مشاهدات گذشته باشد، به صورت خطی یا غیرخطی تعریف می‌شود. به طور کلی مدل‌های سری زمانی شکل‌های مختلفی داشته و فرآیندهای تصادفی مختلفی را معرفی می‌نماید. دو مدل سری زمانی خطی که به طور گسترده در ادبیات مورد استفاده قرار می‌گیرند عبارتند از: مدل خودهمبسته (AR^۴) و مدل میانگین متحرک (MA^۵). از ترکیب این دو مدل، مدل‌های خود همبسته

۱-۳- داده‌های ثبت شده کارت‌بلیت‌های هوشمند با دو بار استفاده در طول شبانه‌روز

در این بخش، تراکنش کارت‌بلیت‌هایی مورد استفاده قرار گرفته که در طول شبانه‌روز فقط دوبار در خط مورد استفاده قرار گرفته‌اند. این دسته از داده‌های AFC بدین صورت گردآوری شده که به ازای هر یک از شماره سریال‌های ثبت شده در هر خط و در هر روز، یک جستجو میان همان داده‌ها انجام می‌گیرد؛ اگر نتیجه این جستجو تنها یک یافته داشته باشد، هر دو تراکنش مربوط به این کارت‌بلیت در این دسته - بندی قرار می‌گیرد. در این دسته از داده‌های AFC که در بخش قبل ایستگاه و زمان پیاده شدن هر تراکنش محاسبه شده بود، می‌توان با استفاده از چند فرض، محل سوار شدن مسافران را به طور نسبتاً مناسب و قابل قبول تخمین زد (شکل ۳). لازم به ذکر است، اگر مسافر تنها در یک جهت رفت یا برگشت سفر خود را با اتوبوس انجام داده و از ایستگاه پیاده شده باشد، سفر دوم خود را از ایستگاهی آغاز می‌کند که مقصد سفر اول بوده ولی برای پیش‌بینی مبدأ سفر اول از توزیع بدست آمده برای ایستگاه‌های سوار شدن مسافران استفاده خواهد شد.

۱-۴- داده‌های ثبت شده کارت‌بلیت‌های هوشمند با بیش از دو بار استفاده در طول شبانه‌روز

در این بخش، تراکنش کارت‌بلیت‌هایی مورد استفاده قرار می‌گیرد که در طول یک شبانه‌روز بیش از دو بار مورد استفاده قرار گرفته‌اند. این دسته از داده‌های AFC بدین صورت گردآوری شده که به ازای هر یک از شماره سریال‌های ثبت شده در هر خط و در هر روز، یک جستجوی میان همان داده‌ها انجام می‌گیرد؛ اگر نتیجه این جستجو بیش از یک یافته داشته باشد، تمام تراکنش مربوط به این کارت‌بلیت در این دسته‌بندی قرار می‌گیرد. در این دسته از داده‌های AFC که در بخش قبل، ایستگاه و زمان پیاده شدن هر تراکنش محاسبه شده بود، می‌توان با استفاده از تعدادی فرض که در شکل ۴ ارائه شده، محل سوار شدن مسافران را به طور نسبتاً مناسب و قابل قبولی تخمین زد. لازم به ذکر است، اگر مسافر تنها در یک جهت رفت یا برگشت سفر خود را با اتوبوس انجام داده

1 Random variable

2 Stochastic Models

3 Artificial Neural Networks

4 Autoregressive (AR)

5 Moving Average (MA)

۲-۲-۳- روش‌های ارزیابی و اعتبارسنجی مدل

برای ارزیابی مدل ساریما از آماره‌ی p -value استفاده شده است. آماره‌ی p -value به منظور تعیین اهمیت هر یک از متغیرهای توضیحی مدل، استفاده می‌شود و بیانگر، میزان احتمال آنکه هیچ ارتباطی بین نمونه و جامعه وجود نداشته باشد، است. در بیشتر مطالعات، مقدار p -value ۵ درصد، به عنوان بیشینه خطای قابل قبول در نظر گرفته می‌شود، اما p -value ۱۰ درصد نیز قابل قبول است [10].

به منظور مقایسه‌ی مدل‌های برآورد شده از آماره R^2 استفاده شد، که نشان‌دهنده‌ی بهبود یا عدم بهبود مدل‌های ساخته شده نسبت به یکدیگر است. در واقع، R^2 بیانگر میزان تغییرات متغیر پاسخ (وابسته) است که توسط مدل قابل توضیح می‌باشد و مقدار آن بین صفر و یک قرار می‌گیرد و بالا بودن آن نشان از توضیح دهندگی بیشتر مدل دارد.

با توجه به اهمیت اساسی پیش‌بینی سری‌های زمانی در بیشتر موقعیت‌های عملی، باید دقت لازم برای انتخاب مدل ویژه بکار گرفته شده و با استفاده از روش‌های مختلف، مدل‌های مختلف مقایسه و دقت پیش‌بینی هر یک تخمین زده شود. در این مطالعه معیار سنجش^{۱۰} مدل‌ها (مدل ساریما و شبکه عصبی پروسپترون چند لایه)، متوسط ریشه مربعات خطا^{۱۱} (رابطه ۲)، متوسط خطای مطلق^{۱۲} (رابطه ۳) و متوسط مطلق درصد خطا^{۱۳} (رابطه ۴) می‌باشد که در آن، $e_t = y_t - f_t$ خطای پیش‌بینی، y_t مقدار واقعی، f_t مقدار پیش‌بینی شده و n اندازه مجموعه آزمایشی است [9].

$$MSE = \sqrt{\frac{1}{n} \sum_{t=1}^n e_t^2} \quad (2)$$

$$MAE = \frac{1}{n} \sum_{t=1}^n |e_t| \quad (3)$$

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{e_t}{y_t} \right| \times 100 \quad (4)$$

میانگین متحرک (ارما^۱) و خود همبسته میانگین متحرک تلفیقی (اریما^۲) ایجاد می‌شود. برای پیش‌بینی سری زمانی‌های فصلی نیز، مدل فصلی خودهمبسته میانگین متحرک تلفیقی (ساریما^۳) استفاده می‌شود [9].

برای تشخیص الگوی اطلاعات سری زمانی ابتدا باید تست ایستا انجام شود. بر این اساس می‌توان، با رسم نمودار خود همبستگی^۴، نمودار خود همبستگی جزئی^۵ و شاخص دیکی-فولر^۶ الگوی اطلاعات از نظر تغییرات ثابت، غیر ثابت (نوسان) و دوره‌ای را تعیین نمود. در این مطالعه با توجه به توزیع ثابت و تغییرات نامنظم در سری زمانی مدل مناسب ساریما، انتخاب شد.

۲-۲-۲- مدل‌های شبکه عصبی

مدل‌های شبکه عصبی مصنوعی روش دیگری برای پیش‌بینی سری‌های زمانی است. یکی از ساده‌ترین و پرکاربردترین شبکه عصبی مصنوعی در پیش‌بینی مسائل، پرسپترون چندلایه^۷ است که از یک لایه پنهان برای پیش بردن شبکه استفاده می‌کند. این مدل با یک شبکه سه لایه‌ای ورودی، پنهان و خروجی مشخص شده است و ممکن است بیش از یک لایه پنهان وجود داشته باشد. گره‌ها در لایه‌های مختلف به عنوان المان‌های پردازشی شناخته می‌شوند [9]. خروجی مدل با استفاده از رابطه ۱ محاسبه می‌شود که در آن y_t خروجی مدل است. y_{t-i} ($i = 1, 2, \dots, p$) ورودی و y_t و اعداد صحیح p و q به ترتیب تعداد ورودی و گره‌های پنهان است. ε_t خطای تصادفی، α_j ($j = 0, 1, 2, \dots, q$) و β_{ij} ($i = 0, 1, 2, \dots, p; j = 0, 1, 2, \dots, q$) وزن‌های ارتباطی

و α_0 و β_{0j} ترم اریب^۸ است.

$$y_t = \alpha_0 + \sum_{j=1}^q \alpha_j g \left(\beta_{0j} + \sum_{i=1}^p \beta_{ij} y_{t-i} \right) + \varepsilon_t, \forall t \quad (1)$$

1 Autoregressive Moving Average (ARMA)

2 Autoregressive Integrated Moving Average (ARIMA)

3 Seasonal Autoregressive Integrated Moving Average (SARIMA)

4 Autocorrelation function (ACF)

5 Partial autocorrelation function (PACF)

6 augmented Dickey-Fuller (ADF)

7 multi-layer perceptrons (MLPs)

8 bias terms

9 R-squared

10 performance metrics

11 Root mean square error (RMSE)

12 The Mean Absolute Error (MAE)

13 The Mean Absolute Percentage Error (MAPE)

جدول ۳. ماتریس مبدأ- مقصد مسافران سالیانه در مسیر پل کریمخان- میدان جمهوری (مسیر رفت)

Total passengers getting on	Destination													Number station
	13	12	11	10	9	8	7	6	5	4	3	2	1	
193874	1518	50527	30770	17642	6836	15580	12628	7987	10655	4333	4613	30785	0	1
82775	457	14461	14125	8554	3857	10869	8648	5584	7146	3780	5294	0	0	2
16037	59	1954	2778	1808	730	2635	1998	1274	1723	1078	0	0	0	3
69121	575	15782	14507	8313	3259	8864	6782	4419	6620	0	0	0	0	4
262770	1889	67841	60055	34656	13942	36610	28752	19025	0	0	0	0	0	5
90045	838	19442	19605	13408	5882	17394	13476	0	0	0	0	0	0	6
103460	1132	25709	22587	18149	9185	26698	0	0	0	0	0	0	0	7
97095	1320	30799	28284	24369	12323	0	0	0	0	0	0	0	0	8
41714	801	16246	13003	11664	0	0	0	0	0	0	0	0	0	9
54040	1100	29524	23416	0	0	0	0	0	0	0	0	0	0	10
45266	1369	43897	0	0	0	0	0	0	0	0	0	0	0	11
430	430	0	0	0	0	0	0	0	0	0	0	0	0	12
0	0	0	0	0	0	0	0	0	0	0	0	0	0	13
1056627	11488	316182	229130	138563	56014	118650	72284	38289	26144	9191	9907	30785	0	Total passengers getting off

Table 3- Annual Origin- destination matrix of bus passengers Karimkhan bridge- Jomhoori square line (went direction)

جدول ۴. ماتریس مبدأ- مقصد مسافران سالیانه در مسیر میدان جمهوری- پل کریمخان (مسیر برگشت)

Total passengers getting on	Destination														Number station
	14	13	12	11	10	9	8	7	6	5	4	3	2	1	
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
231366	6930	3475	534	2784	12153	4137	4804	7420	5884	15034	31242	136969	0	0	2
620105	60610	27706	3312	19322	90165	23319	30902	46089	33398	83935	201347	0	0	0	3
321078	31676	23412	4123	15046	68922	20784	21853	37060	25144	73058	0	0	0	0	4
145554	17892	11866	2196	7956	37604	12230	13323	24434	18053	0	0	0	0	0	5
45060	6291	3786	636	2711	12970	4026	5062	9578	0	0	0	0	0	0	6
64085	10656	7016	1255	4926	22865	7885	9482	0	0	0	0	0	0	0	7
33993	6884	4153	724	2971	14374	4887	0	0	0	0	0	0	0	0	8
19949	4544	2652	470	2061	10222	0	0	0	0	0	0	0	0	0	9
15489	6514	3682	840	4453	0	0	0	0	0	0	0	0	0	0	10
1374	587	547	240	0	0	0	0	0	0	0	0	0	0	0	11
907	350	557	0	0	0	0	0	0	0	0	0	0	0	0	12
3706	3706	0	0	0	0	0	0	0	0	0	0	0	0	0	13
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	14
1502666	156640	88852	14330	62230	269275	77268	85426	124581	82479	172027	232589	136969	0	0	Total passengers getting off

Table 4- Annual Origin- destination matrix of bus passengers on Jomhoori square- Karimkhan bridge line (return direction)

۳- تحلیل اطلاعات

قرار گرفته‌اند. این دسته از داده‌های AFC بدین صورت گردآوری شده که به ازای هر یک از شماره سریال‌های ثبت شده در هر خط و در هر روز، یک جستجو میان همان داده‌ها انجام می‌گیرد؛ اگر نتیجه این جستجو بیش از یک یافته داشته باشد، تمام تراکنش مربوط به این کارت‌بلیت در این دسته-بندی قرار می‌گیرد. در این دسته از داده‌های AFC که در بخش قبل، ایستگاه و زمان پیاده شدن هر تراکنش محاسبه شده بود، می‌توان با استفاده از تعدادی فرض که در شکل ۴(۴) ارائه شده، محل سوار شدن مسافران را به طور نسبتاً مناسب و قابل قبولی تخمین زد. لازم به ذکر است، اگر مسافر تنها در یک جهت رفت یا برگشت سفر خود را با اتوبوس انجام داده و از ایستگاه پیاده شده باشد، سفر خود را از ایستگاهی آغاز می‌کند که در سفر قبل خود، از آن ایستگاه پیاده شده بوده ولی برای پیش‌بینی مبدأ سفر اول از توزیع بدست آمده برای ایستگاه‌های سوار شدن مسافران استفاده خواهد شد.

۴-۱- مدل ساریما (مدل پیش‌بینی تعداد مسافر ایستگاه‌های خط اتوبوس)

به منظور تعیین ایستایی مدل، آزمون دیکی- فولر تعمیم یافته انجام شد که نتایج آن در جدول ۵(۵) ارائه شده و بیانگر ایستایی سری زمانی است. سپس با توجه به ایستایی سری زمانی و تغییرات نامنظم و فصلی آن، همچنین با استفاده از نرم‌افزار استتا^۷ مدل $Sarima(2,0,2)(1,0,2)$ به عنوان بهترین مدل از لحاظ دارا بودن کمترین مقادیر شاخص‌های خطا و بیشترین مقدار ضریب تعیین، انتخاب شد. در این مطالعه، از داده‌های مربوط به تاریخ ۹۴/۳/۱۸ تا ۹۵/۲/۱۸، جهت آموزش مدل استفاده شده و متغیرهای موثر در مدل شامل، شماره ایستگاه، جهت مسیر، ماه، روز هفته و تعطیلی است؛ خروجی مورد انتظار نیز، پیش‌بینی تعداد مسافر ایستگاه‌های هر خط در آن روز می‌باشد. مدل پیشنهادی به صورت رابطه (۵) بوده که در آن:

- i: شمارنده متغیر (= شماره روز)،

در حال حاضر، ۲۲۹ خط^۱ اتوبوس در شهر تهران وجود داشته که شامل ۱۰ خط تندرو و ۲۱۹ خط فیدر است. چگونگی پرداخت کرایه در تمامی خطوط تندرو، هنگام ورود مسافر به ایستگاه و در خطوط فیدر، هنگام پیاده شدن مسافر از اتوبوس است. در ادامه برای نمونه با استفاده از نتایج حاصل از تطبیق اطلاعات AFC و AVL، به تحلیل ماتریس‌های حاصل برای خط فیدر پل کریمخان- میدان جمهوری (مسیر رفت با ۱۳ ایستگاه) و میدان جمهوری- پل کریمخان (مسیر برگشت با ۱۴ ایستگاه) پرداخته شده، لازم به ذکر است تحلیل‌ها بر اساس اطلاعات ۱۳۹۴/۰۲/۱۸ الی ۱۳۹۵/۰۲/۱۸ است [11].

در جدول ۳ ماتریس مبدأ- مقصد مسافران سالیانه در مسیر پل کریمخان به میدان جمهوری نمایش داده شده و همانطور که مشاهده می‌شود، در مجموع ۱۰۵۶۶۲۷ مسافر از تاریخ ۱۳۹۴/۰۲/۱۸ الی ۱۳۹۵/۰۲/۱۸ در این مسیر جابجا شده‌اند. بررسی سالیانه مسافران در مسیر میدان جمهوری به پل کریمخان نیز در جدول ۴ ارائه شده و نشان می‌دهد، از تاریخ ۱۳۹۴/۰۲/۱۸ الی ۱۳۹۵/۰۲/۱۸، در مجموع ۱۵۰۲۶۶۶ مسافر توسط این خط جابه‌جا شده‌اند.

۴- نتایج و بحث

در این بخش، نتایج مدل سری زمانی ساریما و روش شبکه عصبی پروسپترون چند لایه برای برآورد تعداد مسافر ایستگاه‌های خط پل کریمخان- میدان جمهوری ارائه شده و از داده‌های مربوط به مسافران جابجا شده از تاریخ ۹۴/۲/۱۸ تا ۹۵/۲/۱۸ جهت ساخت (آموزش مدل)، تست (آزمون) و اعتبارسنجی مدل استفاده شده است.

۲-۱-۴- داده‌های ثبت شده کارت‌بلیت‌های هوشمند با بیش از دو بار استفاده در طول شبانه‌روز

در این بخش، تراکنش کارت‌بلیت‌هایی مورد استفاده قرار می‌گیرد که در طول یک شبانه‌روز بیش از دو بار مورد استفاده

۱ بر اساس اطلاعات دریافتی ۲۳۰ خط در شهر تهران وجود دارد که وضعیت یکی از خطوط فیدر مشخص نبود و از تحلیل‌های انجام شده حذف گردید.

متغیرهای بکار رفته در مدل $Sarima(2,0,2)(1,0,2)^7$ و نتایج برآورد ضرایب مدل در جدول (۶) ارائه شده است. جهت ارزیابی و آزمون درستی آزمایی مدل از داده مربوط به تاریخ ۹۴/۲/۱۸ تا ۹۴/۳/۱۷ استفاده شده و نتایج معیارهای شاخص خطا برای ارزیابی مدل ساریما برای دو مرحله آموزش و آزمون مدل، در جدول (۷) ارائه شده است.

۴-۲- مدل پرسپترون چند لایه (مدل پیش‌بینی تعداد مسافر ایستگاه‌های خط اتوبوس)

مدلسازی مسافران هر ایستگاه در هر خط اتوبوس با استفاده از شبکه عصبی پرسپترون چند لایه به کمک نرم‌افزار متلب انجام شده است. برنامه به گونه‌ای نوشته شده که قادر است داده‌های دوره ابتدایی را دریافت نموده، عمل نرمال‌سازی داده‌ها را انجام داده و در انتها با استفاده از این داده‌ها تعداد مسافر هر ایستگاه در هر خط اتوبوس را پیش‌بینی نماید. به منظور پیدا نمودن بهترین مدل پیش‌بینی، تعداد ورودی، لایه‌های میانی و گره‌های لایه‌های میانی مدل به برنامه داده شده و شاخص‌های عملکردی بر مبنای هر کدام محاسبه می‌شود. در نهایت شبکه‌ای سه لایه‌ای با ۳۵ لایه ورودی، ۳۵ لایه پنهان (میانی) و ۱ لایه خروجی، به عنوان بهترین مدل انتخاب گردید.

داده‌های مورد استفاده در ساخت این شبکه، مربوط به تعداد مسافر سوار شده در هر ایستگاه در خط پل کریمخان- میدان جمهوری در ۳۶۳ روز از تاریخ ۹۴/۲/۱۸ تا ۹۵/۲/۱۸ بوده و از ۷۰ درصد داده برای آموزش شبکه، از ۱۵ درصد داده‌ها برای ارزیابی شبکه و از ۱۵ درصد باقیمانده جهت آزمون شبکه به صورت تصادفی استفاده شده است. متغیرهای موثر در شبکه شامل، متغیر شماره ایستگاه، جهت مسیر، تاریخ، ماه، متغیر هفته، متغیر تعطیل و تعداد مسافران استفاده کننده از خط پل کریمخان- میدان جمهوری در هر ایستگاه بر اساس روز است. نتایج معیارهای شاخص خطا برای ارزیابی، برای دو مرحله آموزش و آزمون در جدول (۸) ارائه شده است.

- $y_{ds}(i)$: برآورد تعداد مسافر هر ایستگاه خط پل کریمخان- میدان جمهوری در مدل به ازای هر روز،
 - S_s : متغیر ایستگاه، n : متغیر جهت مسیر، m_j : متغیر دوتایی ماه (j شماره ماه)، w_k : متغیر دوتایی روز هفته (k شماره هفته) و h : متغیر دوتایی تعطیلی (۱=تعطیل رسمی)،

- a_d : وایت نویز در سری زمانی با توزیع نرمال با میانگین صفر و انحراف معیار ۲۳۶،

- γ_s : پارامتر ایستگاه، μ : پارامتر تغییر جهت مسیر،
 - α_j : پارامتر متغیر ماه، β_k : پارامتر متغیر روز هفته، δ : پارامتر متغیر تعطیلی،

- φ_{d_i} : پارامتر خودهمبستگی به ازای d درجه تفاوت (۰=) برای هر روز i (i شماره روز)، p درجه خودهمبستگی، P درجه خودهمبستگی فصلی و s دوره تناوب فصلی،

- θ_{d_i} : پارامتر میانگین متحرک به ازای d درجه تفاوت (۰=) برای هر روز i (i شماره روز)، q درجه میانگین متحرک، Q درجه میانگین متحرک فصلی و s دوره تناوب فصلی، و

- d درجه تفاوت (۰=) و D درجه تفاوت فصلی (۰=).

$$y_{ds}(i) = \gamma_s S_s(i) + \mu n(i) + \alpha_j m_j(i) + \beta_k w_k(i) + \delta h(i) + \sum_{l=1}^{p+s} \varphi_{d_l} y_d(i-l) + \sum_{n=s}^{P+p+s} \varphi_{d_p} y_d(i+n) + \sum_{t=P+s} \varphi_{d_{p+p}} y_d(i+t) + a_d(i) + \sum_{o=1}^q \theta_{d_o} a_d(i-o) + \sum_{r=s}^{Q+s} \theta_{d_q} a_d(i+r) + \sum_{f=Q+s} \theta_{d_{q+q}} a_d(i+f) \quad (5)$$

جدول ۵. نتیجه آزمون دیکی فولر برای تعیین ایستایی مدل

Critical values		Dickey Fuller test		Test
-2.570	-2.876	-3.451	-35.855	t test
10%	5%	1%	-	Confidence level

Table 5- Result of the Dickey Fuller test to determine static model

جدول ۶. نتایج برآورد ضرایب مدل هر ایستگاه در خط، $Sarima(2,0,2)(1,0,2)^7$

P-Value	Coefficient	Definition	Variable	Parameter
0.4050	-146.986	-	constant	c
0.0000	685.498	Binary variable: station number 1 to 4=1	Station number 1 to 4	γ_{1-4}
0.0000	882.467	Binary variable: station number 5=1	Station number 5	γ_5
0.1000	374.522	Binary variable: station number 6=1	Station number 6	γ_6
0.1600	259.764	Binary variable: station number 7 to 11=1	Station number 7 to 11	γ_{7-11}
0.0000	-140.180	Binary variable: went direction =1	Direction route	μ
0.0000	-161.080	Binary variable: farvardin=1	Month 1	α_1
0.0010	-130.783	Binary variable: ordibehesht & khordad=1	Month 2 to 3	α_{2-3}
0.0050	-112.577	Binary variable: tir to shahrivar=1	Month 4 to 6	α_{4-6}
0.0000	-65.494	Binary variable: mehr to bahman=1	Month 7 to 11	α_{7-11}
0.0000	349.697	Binary variable: saturday =1	Day 1	β_1
0.0000	364.067	Binary variable: sunday=1	Day 2	β_2
0.0000	371.323	Binary variable: monday=1	Day 3	β_3
0.0000	360.644	Binary variable: tuesday=1	Day 4	β_4
0.0000	337.525	Binary variable: wednesday=1	Day 5	β_5
0.0000	173.140	Binary variable: Thursday=1	Day 6	β_6
0.0000	-256.210	Binary variable: holiday=1	Holiday	δ
0.0000	1.5589	Autoregressive degree=1, Differencing degree=0	Autoregressive parameter 1	φ_{d_1}
0.0000	-0.5667	Autoregressive degree=1, Differencing degree=0	Autoregressive parameter 2	φ_{d_2}
0.0000	0.8774	Seasonal autoregressive degree=1, Time span=7, Seasonal differencing degree=0	Autoregressive parameter 7	φ_{d_7}
0.0000	1.3678	Autoregressive degree=1, Seasonal autoregressive degree=1, Time span=7, Differencing and Seasonal differencing degree=0	Autoregressive parameter 8	φ_{d_8}
0.0000	-0.4972	Autoregressive degree=2, Seasonal autoregressive degree=1, Time span=7, Differencing and Seasonal differencing degree=0	Autoregressive parameter 9	φ_{d_9}
0.0000	-1.1526	Moving Average degree=1, Differencing degree=0	Moving Average Parameter 1	θ_{d_1}
0.0000	0.1954	Moving Average degree=2, Differencing degree=0	Moving Average Parameter 2	θ_{d_2}
0.0000	-0.4913	Seasonal moving Average degree=1, Time span=7, Seasonal differencing degree=0	Moving Average Parameter 7	θ_{d_7}
0.0510	-0.0064	Seasonal moving Average degree=2, Time span=7, Seasonal differencing degree=0	Moving Average Parameter 8	θ_{d_8}
0.0000	0.5662	Moving Average degree=1, Seasonal moving Average degree=1, Time span=7, Differencing and Seasonal differencing degree=0	Moving Average Parameter 9	θ_{d_9}
0.0000	-0.0960	Moving Average degree=2, Seasonal moving Average degree=1, Time span=7, Differencing and Seasonal differencing degree=0	Moving Average Parameter 10	$\theta_{d_{10}}$
0.0000	0.0074	Moving Average degree=1, Seasonal moving Average degree=2, Time span=7, Differencing and Seasonal differencing degree=0	Moving Average Parameter 11	$\theta_{d_{11}}$
0.0000	-0.0013	Moving Average degree=2, Seasonal moving Average degree=2, Time span=7, Differencing and Seasonal differencing degree=0	Moving Average Parameter 12	$\theta_{d_{12}}$

Table 6- Result of bus station passengers' model, $Sarima(2,0,2)(1,0,2)^7$

جدول ۷. مقادیر شاخص‌های خطا برای مدل تعداد مسافر ایستگاه، $Sarima(2,0,2)(1,0,2)^7$

R ²	MAE	MAPE	RMSE	Number of data	Error indicator
0.835	114	99	236	10080	Train model
0.416	237	99	332	54	Test model

Table 7. Values of error indicator for bus station passengers' model, Sarima(2,0,2)(1,0,2)⁷

جدول ۸. مقادیر شاخص‌های خطا برای مدل تعداد مسافر ایستگاه، مدل پرسپترون چند لایه

R ²	MAE	MAPE	RMSE	Number of data	Error indicator
0.946	107	136	186	7069	Train model
0.884	134	280	274	1520	Test model

Table 8. Values of error indicator for bus station passengers' model, multi-layer perceptrons

اطلاعات یکسان بوده و تراکنش ثبت شده از اطلاعات AFC کمترین اختلاف زمانی را با تراکنش ثبت شده از داده‌های AVL داشته باشد، فرض می‌شود که مسافر (صاحب کارت بلیت متناظر با آن تراکنش) در ایستگاهی پیاده شده، که تراکنش پیدا شده با حداقل اختلاف زمانی در آن ایستگاه بوده است. تطبیق اطلاعات AFC و AVL، تنها نشان می‌دهد که مسافران خطوط فیدر شهر تهران، در کدام ایستگاه از اتوبوس پیاده شده‌اند، بنابراین به منظور تخمین محل سوار شدن مسافران اتوبوس، تراکنش کارت بلیت‌های هوشمند ثبت شده به چندین دسته شامل داده‌های ثبت شده بدون شماره سریال کارت بلیت، داده‌های ثبت شده کارت بلیت‌های هوشمند با یکبار استفاده در طول شبانه‌روز، داده‌های ثبت شده کارت بلیت‌های هوشمند با دوبار استفاده در طول شبانه‌روز و داده‌های ثبت شده کارت بلیت‌های هوشمند با بیش از دوبار استفاده در طول شبانه‌روز تقسیم‌بندی شده و برای هر یک از آن‌ها سیاستی جداگانه اجرا گردید. پس از تهیه ماتریس مبدأ-مقصد مسافر، این ماتریس به عنوان پایگاه داده مطالعه مورد نظر قرار گرفته و بر اساس آن مدل ساریما و شبکه عصبی (پرسپترون چند لایه) به منظور پیش‌بینی تعداد مسافر ایستگاه‌های هر خط در روز مشخص پرداخت گردید که نتایج نشان‌دهنده عملکرد بهتر مدل پرسپترون چند لایه می‌باشد.

سپاسگذاری: از مدیریت، مسئولان و کارکنان محترم معاونت مطالعات و برنامه‌ریزی امور زیرساخت و طرح جامع، مرکز مطالعات حمل‌ونقل و برنامه‌ریزی شهر تهران، که حمایت مالی از این مطالعات را برعهده داشتند صمیمانه تشکر می‌شود.

۳-۴- انتخاب مدل مناسب

در این بخش مدل سری زمانی Sarima(2,0,2)(1,0,2)⁷ و شبکه عصبی پرسپترون چند لایه، به عنوان بهترین مدل‌ها برای برآورد تعداد مسافران ایستگاه‌های خط پل کریمخان-جمهوری انتخاب و پرداخت شد و در نهایت تعداد مسافر ایستگاه‌های خط پل کریمخان-میدان جمهوری بر اساس هر دو مدل پیش‌بینی شد. مقایسه نتایج شاخص‌های خطا برای هر دو مدل نشان می‌دهد، مدل پرسپترون چند لایه از لحاظ شاخص‌های خطا در پیش‌بینی تعداد مسافر ایستگاه‌های خط از مدل ساریما برتر است و روش مناسبتری برای برآورد تعداد مسافر ایستگاه‌های خطوط اتوبوس است.

۵- نتیجه‌گیری

امروزه افزایش روزافزون جمعیت، متوسط مالکیت وسیله نقلیه و توسعه ارتباطات بر پیچیدگی‌های سیستم حمل‌ونقل و مشکلات ناشی از آن از جمله تراکم، آلودگی محیط زیست و مصرف منابع تجدیدنپذیر افزوده است. بنابراین تغییر در سیاست‌های حمل‌ونقل شهری و تلاش در جهت توسعه حمل‌ونقل عمومی، به ویژه اتوبوس، یکی از مهم‌ترین اقدامات در حوزه حمل‌ونقل شهری است. در این مطالعه، با استفاده از اطلاعات موجود در سازمان اتوبوسرانی شهر تهران شامل اطلاعات دو سیستم AFC و AVL، مدل برآورد تعداد مسافر ایستگاه‌های اتوبوس شهر تهران تهیه گردید.

بدین منظور، ابتدا باید اطلاعات AFC و AVL مربوط به هر خط اتوبوس مرتب شده و به منظور تعیین ماتریس مبدأ-مقصد مسافر (محل پیاده و سوار شدن مسافر) اطلاعات دو سیستم، تطبیق داده شود. روش کلی بدین صورت است که، وقتی پلاک اتوبوس (کد دستگاه کارت‌خوان اطلاعات AFC و شماره اتوبوس اطلاعات AVL) در هر دو دسته

Public Transport”, *Rotterdam School of Management, Erasmus University*.

[5] Cyprich O., Konecny V. & Kilianova K. 2013 Short-term passenger demand forecasting using univariate time series theory. *Promet, Traffic & Transportation*, **25**(6), 533-541.

[6] Maa Z., Xing J., Mesbah M. & Ferreira L. 2014 Predicting short-term bus passenger demand using a pattern hybrid approach. *Transportation Research Part C*, **39**, 148-163.

[7] Zhoua C., Daib P., Wangc F. & Zhanga Z. 2016 Predicting the passenger demand on bus services for mobile users. *Pervasive and Mobile Computing*, **25**, 48-66.

[8] Frankena M.W. 1978 The demand for urban bus transit in Canada. *Journal of transport economics and policy*. 280-303.

[9] Seyed Abrishami S. E., Amini V., Iran Manesh M. & Mohades Deylami A. 2018 Predicting short-term bus lines passenger demand, case study: Karimkhan bridge- Jomhoori square. *Journal of Transportation Engineering*, Online publish, (In Persian).

[10] Hensher D.A., Rose J.M. & Greene, W.H. 2005 Applied choice analysis. *Published in the United States of America by Cambridge University Press*, New York.

[11] Tehran Bus Company, 2015 Planning and Resource Management Unit (In Persian).

حمایت مالی: این مطالعه با حمایت مالی مرکز مطالعات

حمل و نقل و برنامه‌ریزی شهر تهران انجام شده است.

تعارض منافع: نویسندگان اعلام می‌کنند که هیچ‌گونه

تعارض منافی وجود ندارد.

References

۶- مراجع

[1] Urban Transport Information, 2014 Traffic and Transportation Department Of Tehran Municipality, (In Persian).

[2] Celebi B. B. & Bayraktar D. 2009 Light rail passenger demand forecasting by artificial neural networks. *International Conference on Computers & Industrial Engineering*.

[3] Xue R., Sun D. & Chen, S. 2015 Short-term bus passenger demand prediction based on time series model and interactive multiple model approach. *Discrete Dynamics in Nature and Society*.

[4] Bouman P., Lovric M., Li T., Hurk E., Kroon L. & Vervest P. 2012 Recognizing Demand Patterns from Smart Card Data for Agent-Based Micro-simulation of

Short-term prediction of passenger demand in bus stations, case study: Karimkhan bridge- Jomhoori square

Seyed Ehsan Seyed Abrishami^{1*}, Ali Mohades Deylami², Vajihe Amini³, Maryam Iran Manesh⁴

1-Assis. Professor, Faculty of Civil & Environmental Engineering, Tarbiat Modares University, Tehran, Iran

2- PhD, Faculty of Electronic & Computer Engineering, Tarbiat Modares University, Tehran, Iran

3- MSc, Faculty of Civil & Environmental Engineering, Tarbiat Modares University, Tehran, Iran

4- MSc, Faculty of Civil & Environmental Engineering, Tarbiat Modares University, Tehran, Iran

seyedabrishami@modares.ac.ir

Abstract:

Considering the problems caused by the increasing urban population and the average ownership of the vehicle, changes in urban transport policies and efforts to develop and more use of the public transport, especially the bus, are one of the most important concerns in urban transport planning. A review of various studies suggests that planning for efficient use of bus infrastructures and enhancing the efficiency of public transportation operation in the world, require information on the infrastructure and passenger demand for lines and bus stations. Accordingly, it is necessary to carry out studies to predict passenger demand for bus stations in Tehran. Thus, this study predicts bus stations passenger demand for future short-term periods, using data gathered by AFC (Automated Fare Collection) and AVL (Automatic Vehicle Location) systems. For this purpose, firstly AFC and AVL data was sorted according to the time for each bus line. Since passengers use their smart card while they are getting off the bus it means at the exit station thus identifying their origin station is vital, so that in second step, data of two data bases is compared and matched by writing computer code in Matlab software to determine the origin stations of passengers and then forming origin-destination demand matrix for each bus line in terms of its stations. This matrix is considered as the main data base of the study, a time series analysis, a seasonal autoregressive integrated moving average (SARIMA) and neural network as an artificial model are calibrated based on the available data. Both models' goodness of fit indices are compared in terms of learning and generalization capabilities. For this purpose, initial data is divided into two subsets called learning and test data sets and comparison indices are computed for both aforementioned sets. The models' results show that the multi-layer perceptron neural network model in terms of goodness of fit indices in both learning and generalization capabilities in prediction of bus station passenger demand is better than SARIMA model; however, the manner of influencing different factors such as day of week or month of year in passenger demand in each station is more clear in time series analysis. The passenger demand for each stations in first month in spring is different from the rest months in this season. Months in summer is also show different trends for passenger demand, while all months in fall and the first two months in winter have similar passenger demand in various stations. Official holidays has also significant influence on passenger demand so that reduce passenger demand by approximately 256 persons on average. All days in week have meaningful effects on passenger demand in comparison with Friday so that Monday and Thursday have the highest and the lowest effect on weekday passenger demand in bus stations in comparison with Friday, respectively. This analysis comparison show that if the precision of future prediction is important then neural network outweigh time series regression, while the identification of influential variables on passenger demand is better done by time series analysis.

Keywords: Bus stations passenger demand, Short- term prediction, Times series, Sarima, Neural network, Multi- layer perceptron