

A Novel Method for Recommending Data Plans by Mobile Operators to Maximize Financial Efficiency

Seyed Mohsen

Safavi Koohsareh 

Ph.D. Student in Computer Engineering, Ferdowsi of Mashhad University, Mashhad, Iran

Seyed Amin

Hosseini Sano 

Professor, Engineering Dept., Faculty of Computer Engineering, Ferdowsi of Mashhad University, Mashhad, Iran

Amir Hossein

Mohajerzadeh 

Assistant Professor, Engineering Dept., Faculty of System and Computer Engineering, Sohar University, Oman

Abstract

The primary objective of mobile network operators is arguably to maximize their efficiency. Beyond operational and investment costs, maximizing the utilization of available resources can help them achieve this goal. To this end, operators offer discounted data plans during off-peak hours to encourage users to utilize the network during these times. These data plans are typically based on the average traffic load across the entire network at different times of the day. However, they often overlook the fact that traffic patterns can vary significantly across different population areas within a city at various times. In this paper, different population areas are automatically identified using clustering based on traffic patterns. By identifying these areas and considering the traffic patterns specific to each area, the allocation of appropriate data plans for users, based on the regions they frequent, is analyzed and discussed. Additionally, other potential applications of this clustering method for offering various services are presented, followed by a conclusion.

* Corresponding Author: hosseini@um.ac.ir

How to Cite: Safavi Koohsareh, S.M., Hosseini Sano, S.M., Mohajerzadeh, A.H. (2025). A Novel Method for Recommending Data Plans by Mobile Operators to Maximize Financial Efficiency, *Journal of Business Intelligence Management Studies*, 13(52), 163-188. DOI: 10.22054/ims.2025.81514.2508

1. Introduction

The number of cellular network users and their required bandwidth are continuously increasing (Ericsson, 2022). However, limited wireless frequency bands constrain network capacity, prompting operators to deploy dense base stations to reuse radio frequencies in smaller coverage areas, thereby enhancing capacity. Operators plan for peak usage, leading to base station layouts that often remain underutilized for extended periods, resulting in inefficient use of capital (equipment) and operational (energy and maintenance) costs (Liu et al., 2023). To address this, operators offer discounted plans during low-traffic periods but overlook the varying traffic patterns across urban areas, which could enable tailored offers for different regions. This paper proposes a hierarchical clustering-based method to identify and segment urban areas, design region-specific traffic-based plans, and target appropriate users. The main contribution is improving efficiency by maximizing the utilization of existing cellular networks without expanding capacity, benefiting both operators through increased revenue and users through enhanced satisfaction.

2. Methodology

The best approach to evaluate proposed solutions in cellular networks is to use real-world datasets from mobile operators. Cellular network logs are vast, contain sensitive user and network information, and require algorithms capable of handling large-scale data. In this study, we use a publicly available dataset (Barlacchi et al., 2015) containing telecommunication, weather, news, social media, and power grid data from Milan and Trentino, Italy, spanning November 1, 2013, to June 1, 2014. Our focus is on telecommunication data, specifically Call Detail Records (CDRs), to evaluate the proposed method.

The dataset is processed and analyzed using Python and libraries such as NumPy, Pandas, Scikit-learn, and Matplotlib. The proposed method involves clustering base stations based on traffic patterns, designing region-specific data plans, and targeting users during low-traffic periods.

3.1. Traffic Pattern-Based Region Identification

As mentioned earlier, traffic patterns of cellular base stations vary across urban areas. These patterns are heavily influenced by the stations' locations. For example, base stations in residential areas

exhibit different traffic patterns compared to those in commercial, transportation, or recreational zones (Xu et al., 2017).

Figure 1: Traffic Patterns of Base Stations in Three Different Population Zones

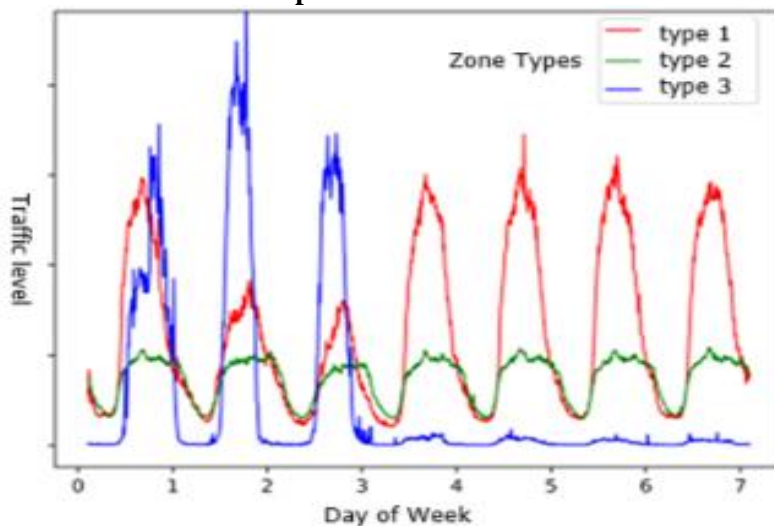


Figure 1 illustrates the traffic patterns of base stations in three different population zones over a week. Zone 3 likely corresponds to recreational areas like amusement parks, with higher traffic on weekends. Zone 1 may represent office areas, with reduced traffic on weekends, while Zone 2 could be industrial or transit areas with consistent traffic throughout the week.

To separate these zones, hierarchical clustering is employed (Abubakar et al., 2022). Instead of using Euclidean distance, which fails to distinguish adjacent zones with different traffic patterns, we use traffic time series as the clustering criterion. The chosen algorithm is agglomerative hierarchical clustering (Kassambara, 2017), as shown in Figure 2. Base stations first remove noise from their data and send average traffic data to a central node every x minutes. At the central node, Euclidean distance is used to measure traffic similarity between stations, reducing dimensionality from two dimensions (time series traffic volume) to one (distance between clusters). Over 80% of time series similarity studies use this metric, though some employ deep learning for feature extraction to improve clustering.

The Euclidean distance between two base stations' traffic time series Q and C is calculated as:

$$D(C, Q) = \sqrt{\sum_{i=1}^n (q_i - c_i)^2}$$

To mitigate sensitivity to variations, preprocessing steps include removing outliers, adjusting offsets, and smoothing noise using moving averages (Keogh & Pazzani, 1998).

The hierarchical clustering dendrogram (Figure 2) determines the optimal number of clusters by identifying the best cut-off line. Two strategies are proposed:

1. Predefine the number of clusters based on comprehensive traffic pattern analysis and use k -means clustering.
2. Use silhouette scoring to dynamically determine the optimal number of clusters based on traffic similarity.

We adopt the second approach, using average silhouette scores (Almeida et al., 2015) to select the optimal number of clusters. This method eliminates the need for predefined cluster counts and provides precise cluster identification.

Once clusters are identified, data plans are designed for each cluster based on their traffic patterns.

3.2. Designing Data Plans

For each cluster, the average traffic profile is calculated, and data plans are designed inversely proportional to traffic volume. The number of offers q in time interval t is determined by:

$$q = \frac{(A - S)}{(A - B)} * (N - C)$$

where A and B are the traffic range bounds, S is the current traffic, N is the maximum number of offers, and C is the minimum (0). Alternative models, such as linear, exponential growth/decay, and logarithmic growth/decay, are also explored (Safavi et al., 2024), as shown in Figures 5 and 6.

3.3. Targeting Users

Users with higher overlap with low-traffic periods are prioritized for data plan offers. A user's average monthly presence in low-traffic intervals is used to rank them. The longest data plans are assigned to users with the highest presence in low-traffic periods, ensuring efficient resource allocation.

4. Results

Simulations represent the method proposed in this paper, utilize 100% of the network's bandwidth capacity.


Results demonstrate the optimal utilization of existing equipment and resources, which directly correlates with increased operator profitability. Those also show that the proposed method can maximize resource efficiency by approximately **40%**, representing the highest possible improvement in network resource utilization.


We can conclude, the proposed method significantly enhances resource utilization and operator profitability by fully leveraging network capacity. While other scenarios improve resource usage to some extent, only the proposed method achieves 100% utilization, highlighting its effectiveness in optimizing network performance and operational efficiency.


Keywords: Mobile Network Operator, Maximizing the Utilization, Cellular Data Plan, Clustering, Traffic Pattern.



روشی برای بیشینه‌سازی بهره‌وری مالی بسته‌های اینترنت اپراتورهای تلفن همراه

سید محسن صفوی کوهساره  دانشجوی دکتری رشته کامپیوتر، دانشگاه فردوسی مشهد، مشهد، ایران

سید امین حسینی سنو  * استاد گروه مهندسی کامپیوتر، دانشگاه فردوسی مشهد، مشهد، ایران

امیر حسین مهاجرزاده  استادیار گروه مهندسی، دانشکده سیستم و کامپیوتر، دانشگاه صحار، عمان

چکیده

شاید بتوان گفت که مهم‌ترین هدف اپراتورهای تلفن همراه، بیشینه کردن بهره‌وری‌شان است. فارغ از هزینه‌های عملیاتی و سرمایه‌گذاری، بهره‌برداری حداکثری از منابع موجود می‌تواند آن‌ها را به این هدف نائل کند. بدین منظور، اپراتورها برای پر کردن ظرفیت‌های خالی شبکه خود در ساعات کم‌بار بودن آن، بسته‌های داده با قیمت‌های مناسب ارائه می‌دهند تا کاربران را برای استفاده از شبکه در این ساعات، تشویق کنند. آن‌ها این بسته‌ها را بر اساس میزان متوسط بار ترافیکی کل شبکه خود در ساعات مختلف روز تنظیم می‌کنند؛ اما به این نکته مهم توجه نمی‌کنند که در نواحی مختلف جمعیتی یک شهر، الگوی ترافیکی در ساعات مختلف متفاوت است. در این مقاله، به صورت اتوماتیک، نواحی مختلف جمعیتی تشخیص داده شده‌اند. این کار به کمک خوشه‌بندی روی الگوی ترافیکی انجام شده است. با تشخیص نواحی جمعیتی و با توجه به الگوی ترافیکی ناحیه مربوطه، برای کاربران، نسبت به منطقه‌ای که بیشتر در آن حضور دارند، بسته مناسب تخصیص می‌یابد. تعداد بسته‌های اختصاص یافته در هر خوشه به کاربران بر اساس زمان کم باری هر یک از خوشه‌ها تعیین می‌شود. در ادامه، با انجام شبیه‌سازی میزان تأثیر روش این مقاله روی افزایش بهره‌وری شبکه اپراتوری نشان داده شده است. در ادامه برای روش ارائه شده در این مقاله، دیگر کاربردهای قابل استفاده برای ارائه سرویس‌های مختلف بیان شده و نهایتاً نتیجه‌گیری انجام گشته است.

کلیدواژه‌ها: اپراتور شبکه موبایل، بیشینه‌سازی بهره‌وری، بسته داده، خوشه‌بندی، الگوی ترافیکی.

مقدمه

تعداد کاربران شبکه‌های سلولی و میزان پهنای باند موردنیاز آن‌ها به‌صورت روزافزون در حال افزایش است (Ericsson, 2022)؛ اما باندهای فرکانسی بی‌سیم محدود هستند و ظرفیت این شبکه‌ها را محدود می‌کنند؛ بنابراین اپراتورها برای افزایش ظرفیت متوسل به استقرار انبوه و متراکم ایستگاه‌های پایه می‌کنند تا از طیف فرکانسی رادیویی در شعاع پوشش دهی کمتر مجدداً استفاده نمایند و این‌گونه ظرفیت شبکه را افزایش می‌دهند.

برنامه‌ریزی اپراتورها برای پیک مصرف انجام می‌شود و بنابراین چینش ایستگاه‌های پایه و تراکم آن‌ها به‌گونه‌ای است که در اوج مصرف پاسخگوی نیازهای مشترکان باشد. مشکل این است که بسیاری از این ایستگاه‌ها در بازه‌های زمانی بزرگی، از کمترین میزان ظرفیت خود استفاده می‌کنند و مابقی ظرفیت آن‌ها خالی می‌ماند (Liu et al., 2023). در نتیجه هم از هزینه‌های سرمایه‌ای^۱ (تجهیزات خریداری و نصب‌شده) و هم از هزینه‌های عملیاتی^۲ (هزینه انرژی و نگهداری تجهیزات) به‌طور بهینه بهره‌برداری نمی‌گردد.

اپراتورها معمولاً برای افزایش بهره‌وری و بهره‌برداری بیشتر از این ظرفیت‌های خالی و کسب درآمد از آن‌ها، طرح‌ها و بسته‌های با قیمت کمتر به کاربران پیشنهاد می‌دهند، اما این طرح‌ها معمولاً برای بازه‌های زمانی است که کل ایستگاه‌های پایه^۳ شبکه در وضعیت کم بار قرار دارند. اپراتورها به این نکته توجه نکرده‌اند که الگوی ترافیکی ایستگاه‌های پایه مناطق مختلف شهری با همدیگر متفاوت است و می‌توان برای کاربران این مناطق، بسته‌ها و طرح‌های متفاوتی پیشنهاد نمود.

در این مقاله، روشی بر مبنای خوشه‌بندی سلسله‌مراتبی پیشنهاد شده است که می‌تواند مناطق مختلف شهری را تشخیص و جدا کند. سپس بر اساس الگوی ترافیکی هر منطقه، بسته‌های پیشنهادی را طراحی می‌کند. در نهایت کاربرانی را که این بسته‌ها باید به آن‌ها پیشنهاد شوند، مشخص می‌کند.

1. CapEx

2. OpEx

3. base station

دستاورد اصلی این مقاله، افزایش بهره‌وری با ارائه روشی برای بهره‌برداری بیشتر از شبکه‌های سلولی موجود بدون افزایش ظرفیت و در نتیجه کسب سود بیشتر توسط اپراتورها و رضایت بیشتر کاربران است.

در ادامه این مقاله، در پیشینه پژوهش‌بخش اول به مرور مقالات و ارائه پیشینه پرداخته شده است. در بخش دوم به انجام جداسازی و تعیین مناطق مختلف شهری پرداخته شده است. سپس در بخش سوم نحوه تعیین بسته‌ها و طرح‌های پیشنهادی به کاربران مورد بحث قرار گرفته است. در ادامه در بخش چهارم انتخاب کاربران هدف برای ارائه پیشنهادها مورد بررسی قرار گرفته است. در نهایت در بخش پنجم به انجام نتیجه‌گیری و بیان راهکارهایی که می‌تواند از این مقاله منتج شود، پرداخته شده است.

پیشینه پژوهش

iDEAL نام روشی است که دانگ و همکاران (Dong et al., 2013) ارائه داده‌اند و پیشنهاد اجاره ظرفیت یا منابع پهنای باند توسط اپراتورها از یک یا چند شرکت ثالث داده است. این گونه در نظر گرفته شده که این شرکت‌ها ایستگاه‌های پایه کوچک^۱ کم‌مصرف را برای تخلیه بار ترافیکی از ایستگاه‌های پایه اپراتور به ایستگاه‌های پایه کوچک خود در ساعات اوج ترافیک برپا و گسترانده‌اند. با این فرض، استراتژی خاموش کردن ایستگاه‌های پایه با تنظیم مجموع هزینه مالی اپراتور شبکه موبایل^۲ و تمام شرکت‌های ثالث به‌عنوان هدف بهینه‌سازی مطرح شده است.

بوسیا و همکاران (Bousia et al., n.d.) این گونه در نظر گرفته‌اند که فروشندگان، ارائه‌دهندگان سلول‌های کوچک^۳ هستند و خریداران اپراتورهای شبکه موبایل که تمایل به خاموش کردن ایستگاه‌های پایه خود برای کاهش هزینه و استفاده از ظرفیت سلول‌های کوچک دارند. مسئله بهینه‌سازی روی این موضوع تعریف شده است که یک مدل مسئله

1. small base station

2. mobile network operator

3. small cell

برنامه‌ریزی صحیح خطی می‌باشد. اپراتورهای شبکه موبایل درخواست اجاره ظرفیت منابع را به صورت جفت پیشنهاد قیمت و ظرفیت مورد نیاز بر اساس پیش‌بینی‌های بار ترافیکی خود، به سلول کوچک مدنظر ارائه می‌کنند و آن‌ها بر اساس سود خود تصمیم خواهند گرفت که به کدام یک از این درخواست‌ها پاسخ مثبت دهند. در واقع مسئله بهینه‌سازی روی بهینه کردن سود اجاره دهندگان سلول‌های کوچک با پر کردن ظرفیت خالی‌شان و هم‌زمان بهینه کردن سود اپراتورها با خاموش کردن ایستگاه‌های پایه خود و استفاده از ظرفیت‌های ارزان‌قیمت‌تر و به‌طور کلی کمینه کردن مصرف انرژی تمرکز دارد.

حسین و همکاران (Hossain et al., 2019)، اشتراک زیرساخت‌ها در مناطق تجمع شبکه‌ها را برای بهینه‌سازی مصرف انرژی مدنظر قرار داده‌اند. در نتیجه یک چهارچوب بهینه از نظر انرژی برای هماهنگی اشتراک ایستگاه‌های پایه بین دو شبکه دسترسی رادیویی سلولی که در یک ناحیه جغرافیایی فعالیت می‌کنند، ارائه شده است. توزیع کاربران و ایستگاه‌های پایه هر دو اپراتور به‌طور مستقل با PPP^۱ مدل شده است. همچنین توان سیگنال دریافتی کاربر مدل شده است. مدل مصرف انرژی به صورت یک مقدار ثابت و یک بخش متغیر با بار ترافیکی تعریف شده است. همچنین رابطه‌ای برای مجموع توان مصرف شده در ایستگاه پایه بعد از توزیع تجهیزات کاربر^۲ میان آن‌ها، تعریف شده است. همچنین رابطه‌ای برای SINR^۳ دریافتی توسط تجهیزات کاربر تعریف شده است.

فنگ و همکاران (Feng et al., 2017a) از معیار کارایی انرژی^۴ برای فرموله کردن مسئله استفاده نموده‌اند. با این تفاوت که گفته شده اگر مستقیماً از معیار کارایی انرژی استفاده کنیم، مجموع نرخ داده کاربران در یک سلول کوچک، حتی اگر تنها تعداد کمی کاربر در سلول کوچک باشند، در سطح بالایی باقی می‌ماند. چون به هر کاربری پهنای باند زیادی اختصاص می‌یابد؛ بنابراین ایستگاه پایه کوچک حتی اگر کاربر کمی داشته باشد، نمی‌تواند خاموش شود. پس به جای استفاده از پهنای باند اختصاص یافته به کاربر، از

-
1. poisson point process
 2. user equipment
 3. signal to interference plus noise ratio
 4. energy efficiency

بدترین حالت پهنای باند اختصاص یافته (یعنی کمترین پهنای باندی که می شود به کاربر اختصاص داد) استفاده کرده است.

در (Feng et al., 2017b) دو راهکار ارائه شده است. در راهکار اول، کاربران لیستی از ایستگاه های پایه را برحسب سرعت مرتب کرده اند و ایستگاه های پایه که در بالای لیست قرار دارند را برای اتصال انتخاب می کنند. در راهکار دوم، هر ایستگاه پایه هزینه مربوط به خود را دارد و کاربران بر اساس تئوری بازی تصمیم می گیرند که به کدام ایستگاه پایه متصل شوند.

چنگ و همکاران (Cheng et al., 2021) سه موضوع اتصال به سلول، خواب سلول^۱ و مسئله تصمیم گیری انگیزشی^۲ برای شبکه سلولی اینترنت اشیا^۳ تغذیه شده به کمک جمع آوری انرژی^۴ با اتصالات دوطرفه^۵ مورد بررسی قرار داده است. آن ها یک بازی Stackelberg را برای بررسی موقعیت میان ایستگاه های پایه فرموله کرده اند.

تمرکز بیشتر مقالات ارائه شده در این زمینه، افزایش بهره وری به کمک کاهش مصرف انرژی، به وسیله خاموش کردن ایستگاه های پایه بدون بار ترافیکی و یا اشتراک آن ها میان چند اپراتور است. نکته قابل توجه این است که با این کار ممکن است هزینه های عملیاتی کاهش یابد، اما بهره وری روی هزینه های سرمایه ای تغییری نخواهد کرد. رابطه روش ارائه شده در این مقاله با روش های پیشنهادی در مقالات بررسی شده این است که در این مقالات به میزان ظرفیت خالی شبکه، ایستگاه های پایه از مدار خارج و خاموش می شوند، اما در روش ما به جای از مدار خارج کردن ایستگاه، سعی می شود از تمام ظرفیت این ایستگاه ها بهره برداری گردد.

-
1. cell sleeping
 2. incentive decision problem
 3. internet of things
 4. energy harvesting aided
 5. full duplex

روش تحقیق

بهترین روش جهت ارزیابی راهکارهای ارائه‌شده در زمینه شبکه‌های سلولی، ارزیابی آن‌ها به کمک مجموعه داده‌های واقعی اپراتورهای تلفن همراه است. حجم داده‌های رخدادننگاری شبکه‌های سلولی بسیار زیاد و حاوی اطلاعات محرمانه کاربران و شبکه می‌باشد و الگوریتم‌های عمل‌کننده در این حوزه، با حجم فراوانی از داده‌ها روبرو هستند. در این مقاله به منظور ارزیابی روش پیشنهادشده، از یکی از محدود مجموعه داده‌هایی که به صورت عمومی منتشر شده و بدین منظور قابل استفاده است، (Barlacchi et al., 2015) استفاده شده است. این مجموعه داده حاوی اطلاعات مربوط به دو شهر میلان و ترنتینو در ایتالیا است که مرکب از اطلاعات مخابراتی، آب‌وهوا، اخبار، شبکه‌های اجتماعی و داده‌های شبکه برق در بازه تاریخ 1 نوامبر 2013 تا 1 ژوئن 2014 می‌باشد. هدف ما استفاده از این مجموعه داده و بهره‌برداری از اطلاعات مخابراتی است که به صورت CDR¹ در پایگاه داده ارائه شده‌اند.

به منظور ارزیابی روش ارائه‌شده، به پردازش و تحلیل این مجموعه داده پرداخته شده است. این کار به کمک زبان برنامه‌نویسی پایتون و کتابخانه‌هایی چون pandas، numpy، sikit-learn و matplotlib انجام شده است.

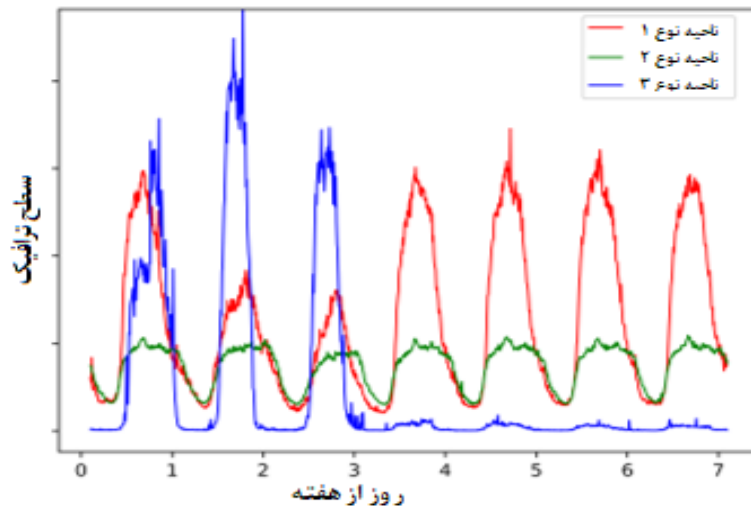
در ادامه مقاله، به تشریح روش ارائه‌شده پرداخته می‌شود.

تشخیص و جداسازی مناطق بر مبنای الگوی ترافیکی

همان‌طور که در مقدمه ذکر شد، الگوی ترافیکی ایستگاه‌های پایه تلفن همراه، در مناطق مختلف شهری متفاوت است. الگوهای ترافیکی ایستگاه‌های پایه می‌تواند به شدت تحت تأثیر موقعیت مکانی‌شان باشند. برای مثال، ایستگاه‌های پایه قرارگرفته در نواحی مسکونی، الگوی ترافیکی متفاوتی نسبت به ایستگاه‌های پایه نواحی تجاری، حمل‌ونقل و تفریحی داشته باشند (Xu et al., 2017).

1. Call detail records

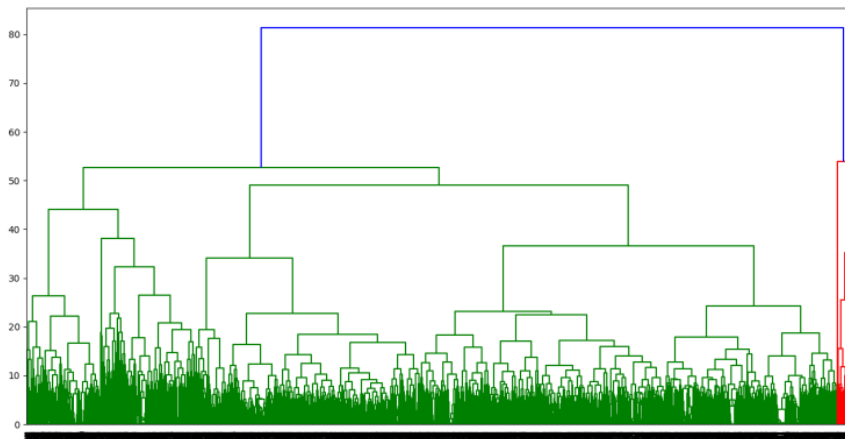
شکل ۱. الگوی ترافیکی ایستگاه‌های پایه مربوط به سه ناحیه جمعیتی مختلف



شکل ۱، الگوی ترافیکی ایستگاه‌های پایه سه ناحیه جمعیتی مختلف را در طول روزهای یک هفته به نمایش گذاشته است. می‌توان حدس زد که ناحیه نوع ۳ احتمالاً به شهرسازی و نواحی تفریحی تعلق دارد که مردم در روزهای آخر هفته از آن استفاده می‌کنند و تجمع جمعیت در این مناطق مربوط به روزهای تعطیل هفته است. ناحیه نوع ۱ نیز می‌تواند مربوط به مناطق اداری باشد که در روزهای تعطیل جمعیت آن‌ها کاهش می‌یابد. ناحیه نوع ۲ نیز می‌تواند مربوط به کارخانجات یا نواحی تردد عمومی است که در تمام ایام هفته جمعیتی یکسان دارند.

شکل ۲. نمودار Dendrogram خوشه‌بندی سلسله مراتبی تجمعی برای لینکیج-کامل با فاصله

همبستگی



برای جداسازی نواحی مختلف جمعیتی از خوشه‌بندی استفاده می‌کنیم (Abubakar et al., 2022). اگر خوشه‌بندی را بر اساس معیار اقلیدسی (فاصله بین ایستگاه‌ها) در نظر بگیریم (Lee et al., 2020)، در صورتی که دو ناحیه مثلاً تجاری و مسکونی به هم چسبیده باشند، قابل تفکیک نخواهند بود؛ بنابراین معیار خود را دنباله زمانی الگوی ترافیکی ایستگاه‌ها انتخاب می‌کنیم. الگوریتم انتخابی ما برای خوشه‌بندی، خوشه‌بندی سلسله‌مراتبی تجمعی^۱ (Kassambara, 2017) است. نمایی از این خوشه‌بندی در شکل ۲ نشان داده شده است. برای اینکه حجم داده‌ها را کمتر کنیم، ایستگاه‌های پایه ابتدا داده‌های نویزی خود را حذف می‌کنند و در هر x دقیقه متوسط ترافیک خود را محاسبه کرده و به گره مرکزی می‌فرستند. در گره مرکزی در انتهای هر روز می‌توان با محاسبه فاصله اقلیدسی که معیاری برای تعیین تشابه دو دنباله زمانی است، تشابه ترافیک ایستگاه‌های پایه را بررسی کنیم (Keogh & Pazzani, 1998). با این کار ما یک کاهش ابعاد نیز از دو بعد (دنباله زمانی حجم ترافیک) به یک بعد (فاصله بین خوشه‌ها) انجام می‌دهیم. در واقع از این الگوریتم به‌عنوان تابع لینکیج در خوشه‌بندی خود استفاده می‌نماییم. قابل ذکر است که

1. agglomerative

بیش از 80% مقالات در زمینه تشخیص مشابهت سری‌های زمانی، از همین معیار استفاده شده است. البته مقالاتی نیز با استخراج ویژگی‌ها با یادگیری عمیق و استفاده از آن‌ها برای خوشه‌بندی، سعی در بهبود هر چه بیشتر خوشه‌بندی داشته‌اند.

اگر سری زمانی حجم ترافیک مربوط به ایستگاه پایه اول را $Q = q_1 \dots q_n$ و سری زمانی حجم ترافیک مربوط به ایستگاه پایه دوم را $C = c_1 \dots c_n$ در نظر بگیریم، فاصله اقلیدسی سری زمانی این دو ایستگاه پایه برابر خواهد بود با

$$D(C, Q) = \sqrt{\sum_{i=1}^n (q_i - c_i)^2} \quad \text{رابطه ۳}$$

چون فاصله اقلیدسی بسیار حساس به تغییرات^۱ است، بنابراین باید قبل از محاسبه آن، با روش‌هایی این تغییرات که در نتایج خوشه‌بندی ما تأثیر منفی می‌گذارند را حذف نماییم. این تغییرات بدین صورت خواهد بود (Keogh & Pazzani, 1998):

تغییر اول: حذف داده‌های خارج از محدوده^۲، تغییر دوم: جابجایی افست^۳ (این تغییر به دلیل این اعمال می‌شود که ترافیک در ایستگاه‌های پایه مختلف قالب خود را حفظ می‌کند ولی ممکن است دامنه‌ی آن تغییر داشته باشد)، تغییر سوم: نویز (عملکرد تابع صاف کننده^۴ بدین گونه است که میانگین هر نقطه داده^۵ را با همسایگانش محاسبه می‌کند. به این عمل میانگین متحرک^۶ هم می‌گویند که در واقع یک فیلتر پایین گذر است).

روش اقلیدسی فوق‌الذکر بر اساس مقایسه شکل ترافیک در طی ۲۴ ساعت خوشه‌بندی را انجام می‌دهد. در خوشه‌بندی سلسله مراتبی، باید تعداد خوشه‌ها را با مشخص کردن بهترین خط برش در نمودار dendrogram مشخص کنیم. این نمودار در شکل ۲ نشان داده شده است. در واقع تعداد خوشه‌ها، تعداد نواحی مختلف از نظر الگوی

1. distortion
2. outlier
3. offset translation
4. smooth
5. data point
6. running average

۱. ابتدا با تحلیل‌های جامع شناختی تعداد مناطق مختلف را از نظر نوع قالب ترافیک تعیین کنیم، سپس به کمک خوشه‌بندی k-means تعداد خوشه‌ها را به طور ثابت برابر این تعداد مناطق در نظر می‌گیریم.
۲. الگوریتمی برای انتخاب بهترین عدد برای تعداد خوشه‌ها، بر اساس میزان فاصله شباهت ایستگاه‌های پایه از هم استفاده کنیم.

شکل ۳: امتیاز سیلوئیت برای تعداد خوشه‌های مختلف



ما راهکار دوم را انتخاب می‌کنیم و از روش متوسط سیلوئیت^۱ (Almeida et al., 2015) برای انجام آن استفاده می‌کنیم. نمایی از این روش در شکل ۳ نشان داده شده است. از مزایای روش دوم این است که برخلاف روش اول که باید تعداد مناطق را به عنوان ورودی به آن می‌دادیم، روش دوم تعداد مناسب انواع مناطق را به طور دقیق به ما اعلام می‌کند. مراحل انجام به صورت زیر خواهد بود.

الگوریتم خوشه‌بندی سلسله‌مراتبی خود را تا رسیدن به یک خوشه واحد ادامه می‌دهیم. حال برای تعداد خوشه‌های $K=1$ تا $K=20$ مراحل ۲ تا ۵ را تکرار می‌کنیم.

1. average silhouette

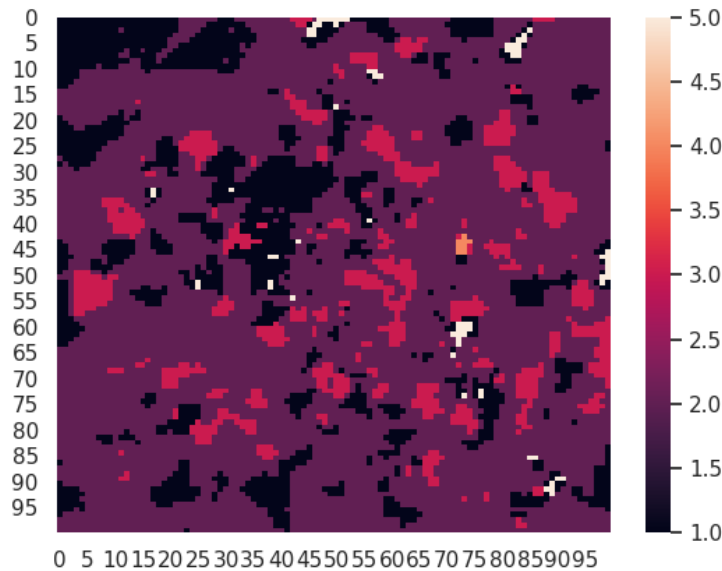
اگر هر ایستگاه پایه را با i مشخص کنیم، متوسط بی شباهتی a_i میان i و تمام دیگر ایستگاه‌های پایه که در یک خوشه با i قرار دارند (C_n) را به صورت $a(i) = \frac{1}{|C_n|-1} \sum_{j \in C_n, i \neq j} d(i, j)$ محاسبه می‌نماییم، که در آن $d(i, j)$ فاصله بین ایستگاه پایه، i و j می‌باشد.

میانگین بی شباهتی i به دیگر خوشه‌ها که i عضو آن‌ها نیست را با b_i نشان می‌دهیم و به صورت میانگین فاصله i از تمام ایستگاه‌های پایه خوشه C ($C \neq C_n$) محاسبه می‌شود. برای هر $i \in C_n$ خواهیم داشت $b(i) = \min_{k \neq n} \frac{1}{|C_k|-1} \sum_{j \in C_k} d(i, j)$. در این رابطه $b(i)$ بدین خاطر است که نزدیک‌ترین خوشه همسایه انتخاب گردد چون این خوشه بهترین گزینه برای عضویت i بعد از خوشه خودش خواهد بود.

در نهایت عرض سیلوئیت i به صورت $s(i) = \frac{b(i)-a(i)}{\max\{a(i), b(i)\}}$ ، $if |C_n| > 1$ و $s(i) = 0, if |C_n| = 1$ بر اساس این تعریف $-1 \leq s(i) \leq 1$ خواهد بود. دقت کنید، هر مقدار $s(i)$ به 1 نزدیک باشد یعنی $b(i) \ll a(i)$ و به عبارتی الحاق i به خوشه C_n شایسته بوده است و بالعکس.

حال مجموع $s(i)$ ها را برای تعداد خوشه‌های k ذخیره می‌کنیم، یعنی $Y[k] = \sum s(i)$
 حال درایه‌های آرایه Y را با هم مقایسه می‌کنیم. ایندکس درایه‌ای که بیشترین مقدار را دارد عدد مناسب تعداد خوشه‌هایمان خواهد بود.

شکل ۴: heatmap نواحی خوشه‌ها بر اساس نوع خوشه



تا به اینجا ما خوشه‌هایی که ترافیک مشابه هم دارند را جدا کردیم که در شکل ۴ قابل مشاهده است.

حال که خوشه‌ها به‌طور کامل مشخص شده و اعضای آن‌ها هم از جهت فضایی و هم الگوی ترافیکی مشابه هم می‌باشند، می‌توان در هر یک از این خوشه‌ها، بسته‌های پیشنهادی را مشخص نمود.

طراحی بسته‌های پیشنهادی

بدین منظور ابتدا متوسط ترافیک خوشه محاسبه می‌شود. با این کار برای هر خوشه، سری زمانی متوسط حجم ترافیک ایستگاه‌های پایه داخل خوشه در ۲۴ ساعت را خواهیم داشت؛ بنابراین ما باید نسبتی معکوس میان این حجم ترافیک و میزان بسته‌های پیشنهادی به کاربران آن خوشه داشته باشیم. با توجه به اینکه فاصله زمانی هر نمونه از الگوی ترافیکی را با نماد t مشخص می‌کنیم (طی بررسی‌های به‌عمل آمده در (Barlacchi et al., 2015) بازه زمانی t برابر ۱۰ دقیقه در نظر گرفته شده است و این مقدار از آمارهای شبکه‌های واقعی

به دست آمده است)، در هر بازه تعداد کاربرانی که پیشنهاد بسته دریافت می کنند، مشخص می کنیم.

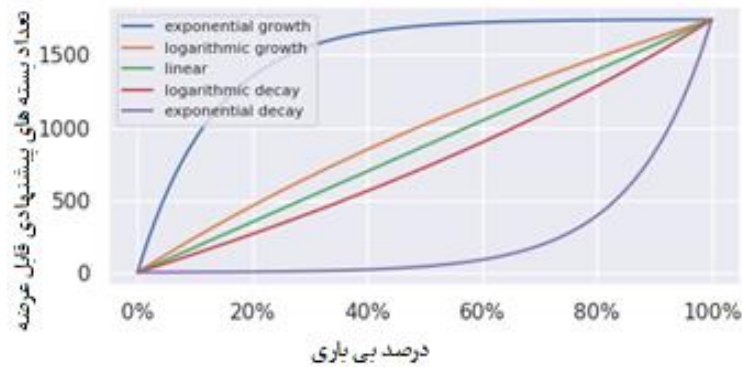
کران بالای تعداد بسته های فعال شونده N در بازه زمانی t برابر تعداد کل کاربران و کران پایین C برابر 0 است. اگر دامنه ترافیک هم $[A, B]$ باشد و همچنین میزان ترافیک فعلی خوشه برابر S باشد، تعداد بسته پیشنهادی به کاربران از رابطه 4 به دست خواهد آمد.

$$q = \frac{(A - S)}{(A - B)} * (N - C) \quad \text{رابطه 4}$$

عدد q به دست آمده نشان دهنده تعداد بسته پیشنهادی قابل عرضه برای دوره زمانی t است.

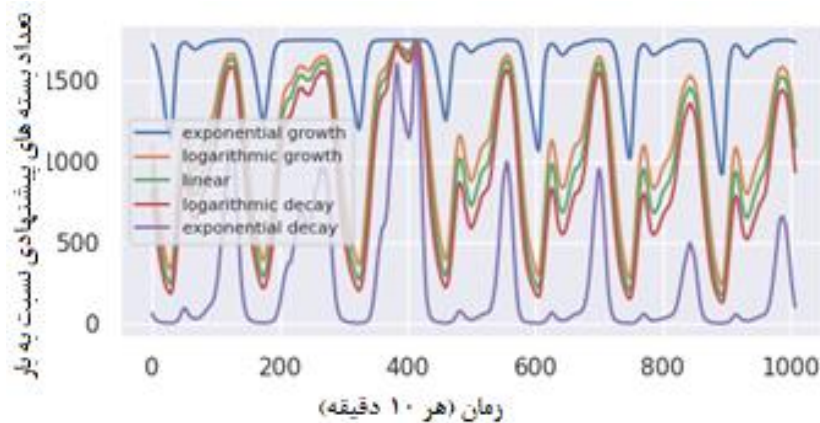
روش فوق نگاشت را به صورت خطی انجام می دهد. تعداد بسته های پیشنهادی را علاوه بر مدل خطی ارائه شده در رابطه 4، **Error! Reference source not found.** می توان با مدل های دیگری نیز محاسبه نمود.

شکل ۵: مدل های مختلف برای محاسبه تعداد بسته های فعال



این مدل ها شامل مدل خطی، رشد نمایی، نزول نمایی، رشد لگاریتمی و نزول لگاریتمی است (Safavi et al., 2024) که در شکل ۵ قابل مشاهده است.

شکل ۶: تعداد بسته‌های قابل ارائه بر اساس ۵ مدل مختلف اعمال شده روی پروفایل ترافیکی خوشه‌ای با ۱۷۴۷ ایستگاه پایه در طی یک هفته



در شکل ۶ نیز تعداد بسته‌های قابل ارائه بر اساس ۵ مدل مختلف ذکر شده، اعمال شده روی خوشه‌ای با ۱۷۴۷ ایستگاه پایه را در طول یک هفته نشان می‌دهد.

با مشخص شدن میزان بسته‌های قابل ارائه در هر بازه زمانی t ، به تعداد بیشترین بازه‌ای که می‌توان از آنها پشت سر هم کنار هم قرار بگیرند و یک بازه زمانی بزرگ‌تر بسازند را انتخاب می‌کنیم و همین‌طور ادامه می‌دهیم تا تمام بسته‌های قابل ارائه تمام شوند. قیمت بسته‌های قابل ارائه نیز با تعداد بسته‌های ارائه شده در بازه زمانی t رابطه عکس و طول زمانی آن رابطه مستقیم خواهد داشت؛ یعنی اگر تعداد بسته‌های قابل ارائه در یک بازه زمانی زیاد باشد، قیمت آن کمتر خواهد شد. همچنین اگر طول بازه زمانی بسته ارائه شده کم باشد، قیمت آن نیز کم خواهد شد.

تعیین کاربران هدف

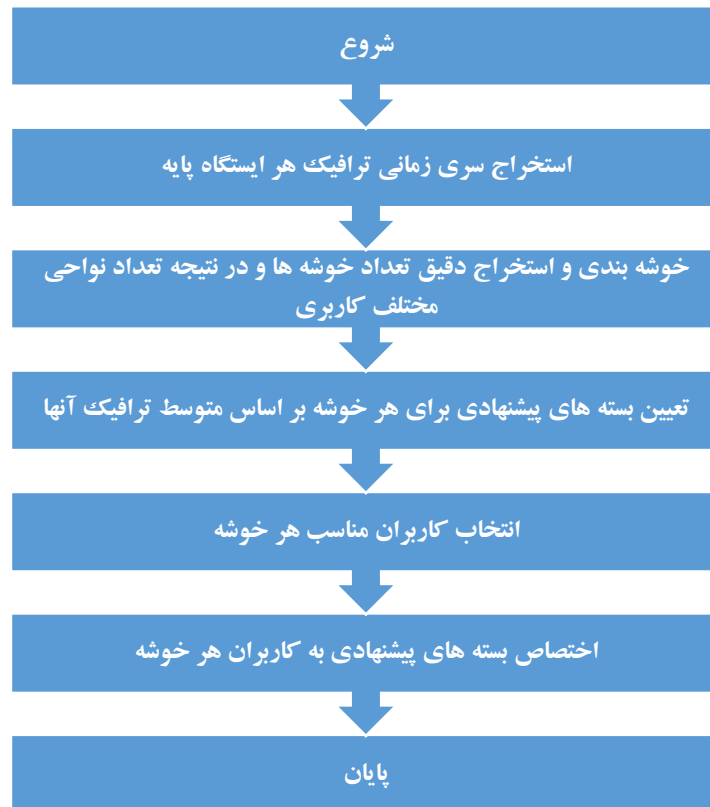
کاربرانی که حضورشان در یک منطقه، همپوشانی بیشتری با زمان کم باری آن منطقه دارد، اولویت انتخاب شدن برای ارائه بسته‌ها هستند.

می‌توان متوسط حضور یک ماهه کاربر را مبنای قرار داد؛ یعنی مجموع اشتراک زمانی حضور کاربر در بازه‌های زمانی کم باری خوشه‌ها را محاسبه کرد؛ بنابراین برای هر خوشه،

کاربران را بر اساس میزان حضورشان در آن خوشه در بازه زمانی کم باری مرتب می‌کنیم و بسته‌های خوشه را نیز بر اساس طول زمانی‌شان مرتب می‌کنیم. حال از بالای لیست به ترتیب طولانی‌ترین بسته را به کاربری که بیشترین زمان را در بازه کم باری خوشه در این خوشه سپری کرده اختصاص می‌دهیم و همین‌طور ادامه می‌دهیم تا تمام بسته‌ها را به کاربران اختصاص دهیم.

الگوریتم کلی روش پیشنهادی به صورت زیر خواهد بود.

شکل ۷: مراحل انجام کار روش پیشنهادی



سناریوهای شبیه‌سازی و مقایسه

در این بخش، بدیل اینکه مقاله‌ای در این زمینه ارائه نشده است، برای انجام شبیه‌سازی و

مقایسه، سناریوهای مختلف را در نظر گرفته و مقایسه را میان این سناریوها انجام می‌دهیم. همان‌طور که در بخش روش تحقیق ذکر شد، ما برای ارزیابی روش ارائه‌شده از مجموعه داده شهر میلان استفاده کرده‌ایم (Barlacchi et al., 2015). مساحت این شهر به مربع‌های با اندازه 235×235 متر تقسیم شده است که مجموعاً 10000 مربع شده است. این کار بر اساس استاندارد WGS84 (EPSG:4326) انجام شده است. از فیلدهای CDR¹ نیز تنها فیلد مصرف اینترنت استفاده شده است.

سناریوها به شرح زیر می‌باشند.

سناریوی اول: هنگامی که هیچ طرحی توسط اپراتور ارائه نشود.

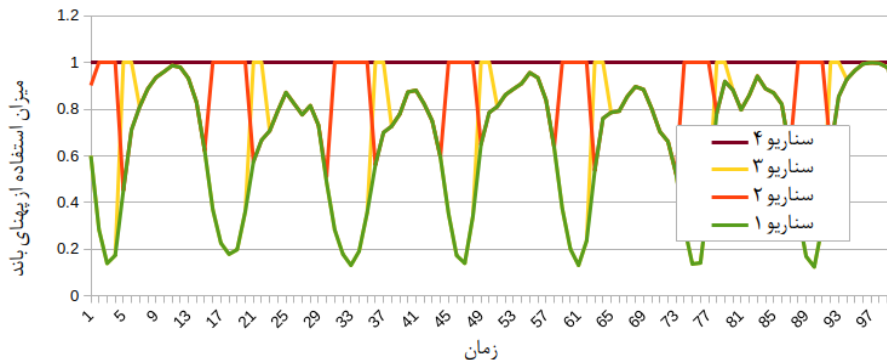
سناریوی دوم: هنگامی که طرح شبانه فعال شود.

سناریوی سوم: هنگامی که طرح صبح فعال شود.

سناریوی چهارم: هنگامی که طرح این مقاله فعال شود.

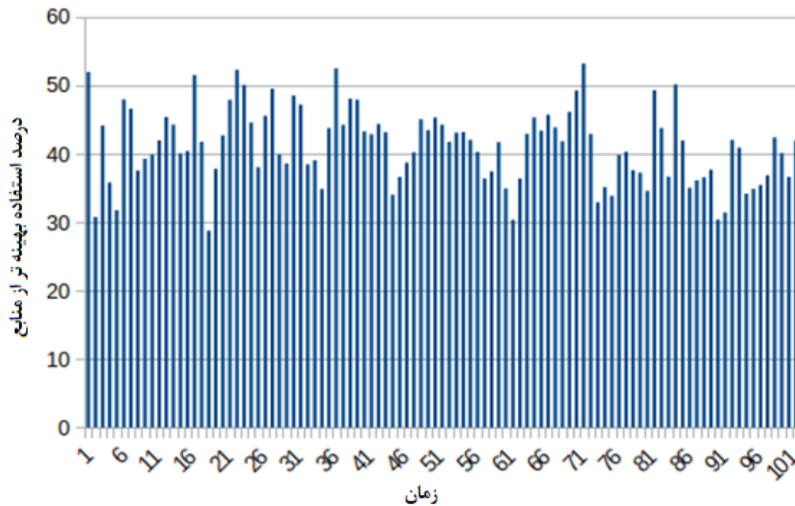
در شبیه‌سازی‌های انجام‌شده، فرض می‌شود که کاربران تمام طرح‌های ارائه‌شده توسط اپراتور را استفاده می‌نمایند. برای انجام شبیه‌سازی، ابتدا خوشه‌بندی را روی مجموعه داده‌ها اعمال می‌کنیم که باعث جداسازی نواحی جمعیتی می‌شود. سپس بسته‌های پیشنهادی را برای هر یک از این خوشه‌ها تعیین می‌کنیم. در نهایت فرض می‌کنیم که تمام بسته‌های پیشنهادی در خوشه در هر سناریو، توسط کاربران فعال و استفاده شده است.

شکل ۸. درصد میزان استفاده بهینه تر از منابع برای مجموع خوشه‌ها (کل شبکه)



شکل ۸، میزان بهره‌برداری از منابع را در حضور سناریوهای مختلف به تصویر کشیده است. همان‌طور که در این شکل قابل مشاهده است، در سناریوی اول، هیچ طرحی ارائه نشده است، بنابراین میزان بهره‌برداری از منابع برابر مجموع ترافیک شبکه عادی روزانه است. در سناریوی دوم که با رنگ قرمز مشخص شده است، طرح شبانه فعال شده است که مجموع بهره‌برداری از منابع در آن برابر مجموع ترافیک شبکه عادی روزانه به اضافه مجموع پهنای باند در دسترس شبانه. سناریوی سوم نیز که با رنگ زرد مشخص شده است، مجموع منابع بهره‌برداری شده در آن برابر ترافیک شبکه عادی به اضافه مجموع پهنای باند در دسترس در بازه صبح. در سناریوی چهارم که روش ارائه شده توسط این مقاله است، از تمام ظرفیت پهنای باند شبکه به صورت صددرصدی بهره‌برداری شده است. واضح است که هیچ سناریویی به جز سناریوی چهارم از تمام ظرفیت شبکه استفاده نمی‌نماید. اگر بخواهیم، میزان مصرف بهینه‌تر از تجهیزات و منابع موجود را به تصویر بکشیم، نمودار آن به شرح شکل ۹ خواهد بود. این مقدار رابطه مستقیمی با افزایش میزان سود حاصله اپراتور خواهد داشت.

شکل ۹: درصد میزان استفاده بهینه‌تر از منابع برای مجموع خوشه‌ها (کل شبکه)



شکل ۹ نشان می‌دهد که بهره‌وری از منابع موجود شبکه را می‌توان با روش ارائه‌شده توسط این مقاله، در بیشینه‌ترین حالت، حدود ۴۰ درصد افزایش داد.

نتیجه‌گیری و کارهای آینده

در این مقاله به ارائه روشی پرداخته شد که مصرف منابع اپراتور را در زمان‌های بلااستفاده بودن به اوج خود می‌رساند؛ بنابراین اپراتور می‌تواند با بهره‌گیری از تمام توان، به بیشینه سود حاصله و بهره‌وری بالا دست یابد. این کار به کمک خوشه‌بندی روی الگوی ترافیکی انجام می‌پذیرد. با تشخیص نواحی جمعیتی از طریق خوشه‌بندی و با توجه به الگوی ترافیکی هر خوشه، برای کاربران، نسبت به منطقه‌ای که بیشتر در آن حضور دارند، بسته مناسب تخصیص می‌یابد. تعداد بسته‌های قابل ارائه و طول زمانی بسته‌ها نیز در هر خوشه محاسبه می‌شود.

در این کار تمرکز بر بیشینه‌سازی بهره‌وری اپراتور بود، اما در کارهای آینده می‌توان این کار را از منظر کاربری که بسته‌های پیشنهادی به او اختصاص می‌یابد، بیشتر موردبررسی قرار داد. مثلاً چه درصدی از کاربران بسته پیشنهادی را فعال می‌کنند؛ بنابراین

می توان این پارامتر را در تصمیم گیری در مورد قیمت و دیگر پارامترهای بسته ها مدنظر قرار داد.

همچنین ما مناطق خوشه ها را از منظر جامع شناختی مورد بررسی قرار ندادیم. مثلاً اگر سطح بودجه و درآمد افراد حاضر در خوشه بالا یا پایین باشد، می تواند در مورد طول بسته های پیشنهادی، قیمت، سرعت و حجم ترافیک آن ها اثرگذار باشد. ما در این کار از تحلیل ترافیک ایستگاه پایه برای پیشنهاد بسته ها توسط اپراتورها استفاده کردیم؛ اما می توان از این تحلیل و روش خوشه بندی برای ارائه خدمات اجتماعی مانند زمان ارائه خدمات برای مناطق مختلف شهر استفاده کرد. مثلاً زمان جمع آوری زباله ها، زمان انجام تعمیرات لوله کشی، زمان فعالیت پارکینگ های عمومی، تصمیم گیری ها در زمینه کنترل ترافیک، زمان بندی خدمات حمل و نقل، توسعه زیرساخت ها و ...

سپاسگزاری


از دانشگاه فردوسی مشهد به عنوان حامی مادی و معنوی این پژوهش تقدیر و تشکر می شود.


تعارض منافع


تعارض منافع وجود ندارد.

ORCID

Seyed Mohsen Safavi
Koohsareh
Seyed Amin Hosseini
Sano
Amirhossein
Mohajerzadeh

 <https://orcid.org/0009-0009-1271-1225>

 <https://orcid.org/0000-0002-0838-1800>

 <https://orcid.org/0000-0002-2630-0429>

References

1. Abubakar, A. I., Mollel, M. S., Ozturk, M., Hussain, S., & Imran, M. A. (2022). A lightweight cell switching and traffic offloading scheme for energy optimization in ultra-dense heterogeneous networks. *Physical Communication*, 52, 101643.
2. Almeida, J. P., Oliveira, J. F., & Pinto, A. A. (2015). *Operational Research: IO 2013 - XVI Congress of APDIO, Bragança, Portugal, June 3-5, 2013*. Springer International Publishing. <https://books.google.com/books?id=TpyoCgAAQBAJ>
3. Barlacchi, G., De Nadai, M., Larcher, R., Casella, A., Chitic, C., Torrisi, G., Antonelli, F., Vespignani, A., Pentland, A., & Lepri, B. (2015). A multi-source dataset of urban life in the city of Milan and the Province of Trentino. *Scientific Data*, 2(1), 150055. <https://doi.org/10.1038/sdata.2015.55>
4. Bousia, A., Kartsakli, E., Member, S., Antonopoulos, A., Member, S., Alonso, L., Member, S., Verikoukis, C., Member, S., & Motivation, A. (n.d.). *Multiobjective Auction-Based Switching-Off Scheme in Heterogeneous Networks To Bid or Not to Bid*.
5. Bousia, A., Kartsakli, E., Member, S., Antonopoulos, A., Member, S., Alonso, L., Member, S., Verikoukis, C., Member, S., Motivation, A., Antonopoulos, A., Member, S., Alonso, L., Member, S., Verikoukis, C., Member, S., & Motivation, A. (2016). Multiobjective Auction-Based Switching-Off Scheme in Heterogeneous Networks To Bid or Not to Bid. *IEEE Transactions on Vehicular Technology*, 65(11), 9168–9180. <https://doi.org/10.1109/TVT.2016.2517698>
6. Cheng, Y., Zhang, J., Zhang, J., Zhao, H., Yang, L., & Zhu, H. (2021). Small-Cell Sleeping and Association for Energy-Harvesting-Aided Cellular IoT With Full-Duplex Self-Backhubs: A Game-Theoretic Approach. *IEEE Internet of Things Journal*, 9(3), 2304–2318.
7. Dong, W., Rallapalli, S., Jana, R., Qiu, L., Ramakrishnan, K. K., Razoumov, L., Zhang, Y., & Cho, T. W. (2013). iDEAL: Incentivized dynamic cellular offloading via auctions. *IEEE/ACM Transactions on Networking*, 22(4), 1271–1284.
8. Ericsson. (2022). Ericsson Mobility Report. *Ericsson, November*, 40. www.ericsson.com/mobility-report
9. Feng, M., Mao, S., & Jiang, T. (2017a). BOOST: Base station on-off switching strategy for green massive MIMO HetNets. *IEEE Transactions on Wireless Communications*, 16(11), 7319–7332. <https://doi.org/10.1109/TWC.2017.2746689>
10. Feng, M., Mao, S., & Jiang, T. (2017b). BOOST: Base station on-off switching strategy for green massive MIMO HetNets. *IEEE Transactions on Wireless Communications*, 16(11), 7319–7332. <https://doi.org/10.1109/TWC.2017.2746689>
11. Hossain, F., Munasinghe, K. S., & Jamalipour, A. (2019). Energy-efficient inter-RAN cooperation for non-collocated cell sites with base station selection and user association policies. *Wireless Networks*, 25, 269–285. <https://doi.org/10.1007/s11276-017-1556-4>

12. Kassambara, A. (2017). Multivariate Analysis I: Practical Guide To Cluster Analysis in R. Unsupervised Machine Learning. In *Taylor & Francis Group*. STHDA. <https://books.google.com/books?id=plEyDwAAQBAJ>
13. Keogh, E. J., & Pazzani, M. J. (1998). An Enhanced Representation of Time Series Which Allows Fast and Accurate Classification, Clustering and Relevance Feedback. *Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining*, 239–243. <http://dl.acm.org/citation.cfm?id=3000292.3000335>
14. Lee, W., Jung, B. C., & Lee, H. (2020). DeCoNet: Density clustering-based base station control for energy-efficient cellular IoT networks. *IEEE Access*, 8, 120881–120891.
15. Liu, Z., Chen, X., Yang, Y., Chan, K. Y., & Yuan, Y. (2023). Joint cell zooming and sleeping strategy in ultra dense heterogeneous networks. *Computer Networks*, 220, 109482.
16. Safavi, S. M., Seno, S. A. H., & Mohajerzadeh, A. (2024). An Adaptive Cell Switch Off framework to Increase Energy Efficiency in Cellular Networks. *Wireless Personal Communications*, 135(4), 2011–2037. <https://doi.org/10.1007/s11277-024-11027-0>
17. Xu, F., Li, Y., Member, S., Wang, H., Zhang, P., & Jin, D. (2017). Understanding Mobile Traffic Patterns of Large Scale Cellular Towers in Urban Environment. *IEEE/ACM TRANSACTIONS ON NETWORKING*, 25(2), 1147–1161.

استناد به این مقاله: صفوی کوهساره، سید محسن، حسینی سنو، سید امین، مهاجرزاده، امیرحسین. (۱۴۰۴). روشی برای پیشینه‌سازی بهره‌وری مالی بسته‌های اینترنت اپراتورهای تلفن همراه، *مطالعات مدیریت کسب و کار هوشمند*، ۱۳(۵۲)، ۱۶۳–۱۸۸. DOI: 10.22054/ims.2025.81514.2508



Journal of Business Intelligence Management Studies is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License..