Industrial Management Journal

Print ISSN: 2008-5885

Online ISSN: 2423-5369

Identifying the Latitude and Longitude of ATMs in ATM Networks

Niloofar Haghjoo

Ph.D., Department of Bioinformatics, University of Tehran, Tehran, Iran. E-mail: n.haghjoo@dotin.ir

Mohammad Rahmati

MSc., Department of Computer Engineering, Faculty of Computer Engineering, University of Isfahan, Isfahan, Iran. E-mail: mo.rahmati@dotin.ir

Ali Zare Mirakabad *

*Corresponding Author, Lecture, Department of Management, College of Management, University of Tehran, Tehran, Iran. E-mail: alizare@ut.ac.ir

Abstract

Objective

The geographical positioning of Automated Teller Machines (ATMs) is a pivotal data point that significantly aids in the analytical process and decision-making for a multitude of critical banking and economic determinations. Given the constraints imposed by the insular viewpoint prevalent in the nation's banking ecosystem, maintaining a consolidated perspective of all ATMs' geographical locations is not feasible. In this study, we utilized the ATM Location Prediction (ATMLP) algorithm to determine these machines' geographical coordinates. This data is indispensable and plays a cardinal role in the implementation of a multitude of artificial intelligence algorithms.

Methods

The ATMLP algorithm comprises three primary stages. The first stage involves constructing a bipartite user-location graph. The relationship between users is derived from transactional interactions, while the relationship between geographical locations is established using devices with known locations. The second stage involves the computation of two crucial indices: spatial similarity and neighborhood similarity, within the ATM network using the bipartite graph. This stage also includes a time-space distance finding module, which has two steps in its procedure: finding co-located ATMs and then clustering them. Distance-based features are assigned to edges because they reflect the similarity level between the pair of ATMs, nodes connected by edges. The third stage of the algorithm fine-tunes the results for better accuracy. In this process, low-confidence edges are filtered out by leveraging similarity metrics from the previous stage and cosine similarity between pairs of ATMs. In the end, the algorithm reports the geographical

latitude and longitude for each ATM, plus the probability score indicating how correct it is.

Results

By leveraging 2100 ATM locations (a portion of the data available in Datis Arian Qeshm Company) and examining 562609790 transactions in four months from the start of April 2022 to the end of July 2022, we identified the location of 4000 existing ATMs across the country belonging to 12 banks. The results obtained indicate a high credibility of the algorithm (80.95%).

Conclusion

In this study, we applied a developed method in banking to predict edges in location-based social networks, and using it, we accurately estimated the geographical coordinates - latitude and longitude - of ATMs on a national scale. Location-based social networks, due to data integration at multiple levels, enable problem-solving that was previously impossible. The use of these methods has less processing cost and higher speed due to the use of algorithms and graph-based databases, and they provide more accurate results**.** This study has significant implications for future research in banking technology, particularly about location prediction for ATMs.

Keywords: ATM location prediction algorithm, ATMs, Link prediction, Location-based social networks.

Citation: Haghjoo, Niloofar; Rahmati, Mohammad & Zare Mirakabad, Ali (2024). Identifying the Latitude and Longitude of ATMs in ATM Networks. *Industrial Management Journal*, 16(2), 282-302. (*in Persian*)

Industrial Management Journal, 2024, Vol. 16, No 2, pp. 282- 302
Published by University of Tehran, Faculty of Management
https://doi.org/10.22059/IMJ.2024.366690.1008100
Article Type: Research Paper
© Authors
Received: October 14, 2023
Received in revised form: January 13, 2024
Accepted: April 24, 2024
Published online:July 20, 2024

مدیریت صنعتی
شاپا چاپی: ۵۸۸۵-۲۰۰۸
شاپا الکترونیکی: ۵۳۶۹-۲۴۲۳

ارائه مدلی به‌منظور مکان‌یابی دستگاه‌های خودپرداز بر اساس تراکنش‌ها در شبکه دستگاه‌های خودپرداز

نیلوفر حق‌جو
دکتری، گروه بیوانفورماتیک، دانشکده بیوانفورماتیک، دانشگاه تهران، تهران، ایران. رایانامه: n.haghjoo@dotin.ir

محمد رحمتی
کارشناس ارشد، گروه مهندسی کامپیوتر، دانشکده مهندسی کامپیوتر، دانشگاه اصفهان، اصفهان، ایران. رایانامه: mo.rahmati@dotin.ir

علی زارع *
* نویسنده مسئول، مربی، گروه مدیریت، دانشکدگان مدیریت، دانشگاه تهران، تهران، ایران. رایانامه: alizare@ut.ac.ir

چکیده

هدف: مکان جغرافیایی دستگاه‌های خودپرداز، داده‌ای کلیدی است که تجزیه‌وتحلیل آن برای پاسخ به بسیاری از تصمیم‌های مهم بانکی و اقتصادی راه‌گشاست. با توجه به محدودیت‌های ناشی از نگرش بسته در اکوسیستم بانکی کشور، امکان در دست داشتن یکپارچهٔ مکان‌های جغرافیایی تمامی دستگاه‌های خودپرداز میسر نیست. در مطالعهٔ حاضر که از نوع کاربردی با روش توصیفی – هم‌بستگی است به‌کمک داده‌های موجود در شرکت داتیس آرین قشم (داتین)، طول و عرض جغرافیایی این دستگاه‌ها با استفاده از الگوریتم پیش‌بینی مکان دستگاه‌های خودپرداز به‌دست آمده است؛ زیرا این داده‌ها در پیاده‌سازی بسیاری از الگوریتم‌های هوش مصنوعی لازم و ضروری است و نقش اساسی ایفا می‌کند.

روش: الگوریتم مکان‌یابی ارائه شده از سه مرحله کلی تشکیل شده است. در این الگوریتم، ابتدا گراف دوبخشی کاربر ـ مکان تشکیل می‌شود. ارتباط بین کاربران با استفاده از تراکنش‌هایی که کاربران انجام داده‌اند، استخراج می‌شود و ارتباط بین مکان‌های جغرافیایی با استفاده از دستگاه‌هایی که مکان معلومی دارند، تشکیل می‌شود. در مرحله بعد با استفاده از گرافِ دو بخشیِ تشکیل شده، دو شاخص شباهت مکانی و شباهت همسایگی در شبکه خودپردازها محاسبه می‌شود. در همین مرحله با استفاده از ماژول یافتن فاصلهٔ زمانی ـ مکانی که خود شامل دو مرحله یافتن خودپردازهای هم‌مکان و خوشه‌بندی آن‌هاست، اجرا می‌شود و ویژگی‌هایی به یال‌ها اختصاص داده می‌شود که بر فاصله به همراه میزان شباهتِ دو خودپرداز (نودهای متصل‌کنندهٔ یال‌ها) مبتنی است. مرحله سوم در این الگوریتم برای افزایش دقت نتایج طراحی شده است و شامل فیلتر کردن یال‌هایی است که با اطمینان پایین با استفاده از شباهت به‌دست‌آمده از مرحله قبل حاصل شده است و شباهت کسینوسی دو دستگاه خودپرداز است. در نهایت با استفاده از یال‌ها و دقت به‌دست‌آمده برای هر خودپرداز، طول و عرض جغرافیایی به همراه احتمال درستی گزارش می‌شود.

یافته‌ها: با استفاده از محل استقرار ۲۱۰۰ خودپرداز (بخشی از داده‌های موجود در شرکت داتیس آرین قشم) و بررسی ۵۶۲۶۰۹۷۹۰ تراکنش در بازه زمانی چهار ماهه، از ابتدای فروردین ماه سال ۱۴۰۱ تا پایان تیرماه همان سال، محل ۴۲۰ خودپردازِ موجود در کل کشور شناسایی شد. نتایج به‌دست‌آمده نشان‌دهنده اعتبار عالی الگوریتم (۸۰/۹۵درصد) است.

نتیجه‌گیری: در این مطالعه با کاربست روش توسعه داده شده در حوزه بانکداری، به پیش‌بینی یال در شبکه‌های اجتماعی مبتنی بر مکان پرداخته شد و با استفاده از آن، طول و عرض جغرافیایی دستگاه‌های خودپرداز در سطح کشور تخمین زده شد. شبکه‌های اجتماعی مبتنی بر مکان، به‌دلیل ادغام داده‌ها در چندین سطح، حل مسائلی را ممکن می‌سازند که تا پیش از این امکان‌پذیر نبود. استفاده از این روش‌ها، به‌خاطر استفاده از الگوریتم‌ها و پایگاده داده مبتنی بر گراف، هزینه پردازشی کمتر و سرعت بیشتری دارند و نتایج دقیق‌تری را ارائه می‌دهند.

کلیدواژه‌ها: الگوریتم پیش‌بینی مکان دستگاه‌های خودپرداز، پیش‌بینی یال، دستگاه‌های خودپرداز، شبکه‌های مبتنی بر مکان.

استناد: حق‌جو، نیلوفر؛ رحمتی، محمد و زارع، علی (۱۴۰۳). ارائه مدلی به‌منظور مکان‌یابی دستگاه‌های خودپرداز بر اساس تراکنش‌ها در شبکه دستگاه‌های خودپرداز. *مدیریت صنعتی*، ۱۶(۲)، ۲۸۲– ۳۰۲.

مدیریت صنعتی، ۱۴۰۳، دوره ۱۶، شماره ۲، صص. ۲۸۲– ۳۰۲
ناشر: دانشکدهٔ مدیریت دانشگاه تهران
نوع مقاله: علمی پژوهشی
© نویسندگان
تاریخ دریافت: ۱۴۰۲/۰۷/۲۲
تاریخ ویرایش: ۱۴۰۲/۱۰/۲۳
تاریخ پذیرش: ۱۴۰۳/۰۲/۰۵
تاریخ انتشار: ۱۴۰۳/۰۴/۳۰
doi: https://doi.org/10.22059/IMJ.2024.366690.1008100

مقدمه

در بسیاری از صنایع، رقابت یکی از موضوعات اجتناب‌ناپذیر است. رقابت می‌تواند با ابزارهای متعددی شکل بگیرد که یکی از این رویکردها، رقابت مبتنی بر دانش است. فرانسیس بیکن، نخستین بار عبارت «دانش، قدرت است» را نقل کرد (به نقل از انگرمن[1]، ۲۰۰۷). امروزه با وجود تعاریف گوناگون از دانش، بسیارند افرادی که نه تنها با این عبارت موافق هستند، بلکه آن را تجربه و درک کرده‌اند. لازم است همین ابتدا تفاوت اساسی بین اطلاعات و دانش را مد نظر قرار دهیم. در حال حاضر، اطلاعات انبوهی در حال جمع‌آوری هستند؛ اما از این اطلاعات به‌صورت خام نمی‌توان استفادهٔ درخور توجهی کرد. با استفاده از روش‌های تحلیل داده یا داده‌کاوی، می‌توان اطلاعات را به دانش و پس از آن، به بصیرت تبدیل کرد (لی و کتینگر[2]، ۲۰۲۲).

بنابراین نخستین مرحله برای ارائهٔ خدمات رقابتی در هر حوزه‌ای، جمع‌آوری داده است. در سیستم بانکداری، خدمات متنوعی به مشتریان ارائه می‌شود. برخی از این سرویس‌ها مانند خدمات الکترونیکی و تنوع آن، در این رقابت سهم بسزایی دارند که لازمه توسعهٔ بانک‌ها، توسعهٔ مداوم این خدمات است (میرفخرالدینی و امیری، ۱۳۸۹). یکی از خدماتی که به تحول در این حوزه منجر شد، ظهور دستگاه‌های خودپرداز[3]، بدون نیاز به نیروی انسانی است. سالانه تعداد زیادی تراکنش (حدود ۱۲ میلیارد تراکنش در سال) از طریق این دستگاه‌ها انجام می‌شود. یکی از اطلاعات استراتژیک برای بانک‌ها و سایر مؤسسه‌های مالی، محل استقرار تمامی خودپردازها (شامل خودپردازهای دیگر بانک‌ها) است. برای مثال، در کارهای مشابه (آقا قلیزاده، سیار و معتدل، ۱۳۹۸؛ نظری گنجه و میرزاپور آل هاشم[4]، ۲۰۲۰؛ گورگول و سودر[5]، ۲۰۱۸) ، نواحی مناسب برای قرار دادن خودپردازهای جدید با استفاده از تاریخچهٔ مکان خودپردازها و در مطالعه‌ای دیگر (سوریا، کاندیازا و دانتس[6]، ۲۰۲۲) میزان بهره‌وری هر خودپرداز در هر ناحیه با استفاده از چندین معیار، از جمله فاصله خودپردازهای موجود با خودپردازهای رقیب، به‌دست آمده است. این اطلاعات در جایابی محل دستگاه‌ها برای کسب درآمد بیشتر بسیار مؤثر است؛ زیرا از یک سو سبب می‌شود که دستگاه در نقطه‌ای قرار گیرد که تراکنش‌های زیادی داشته باشد و پیش از سایر رقبا، نقطه‌ای بهینه از نظر هزینه و درآمد را پیدا کرد و از سوی دیگر، می‌توان هزینهٔ لازم برای بازاریابی و جابه‌جایی دستگاه‌های خودپرداز را کاهش داد. همچنین در مقاله‌های متعدد دیگری مانند (تاکینوا و گولیوا[7]، ۲۰۲۳؛ راستوگی، شارما، موخرجی و باغل[8]، ۲۰۲۳) بر لزوم مکان‌یابی بهینه برای دستگاه‌های خودپرداز تأکید شده است. در یکی از مطالعات (آدسینا، آدوویی، مورنیکجی، اوگوندل و باباتونده[9]، ۲۰۲۲) مکان فعلی دستگاه‌های خودپرداز منطقه‌ای مشخص، از منظر بهینه بودن در پوشش‌دهی بررسی شده و بر اهمیت انتخاب مکان مناسب برای خودپردازهای بعدی تأکید کرده است. در برخی دیگر از مطالعات، اهمیت موقعیت جغرافیایی، بر حفظ امنیت کاربران

1. Engerman
2. Li & Kettinger
3. Automated Teller machine Location Prediction (ATMLP)
4. Nazari-Ganje & Mirzapour Al-E Hashem
5. Gurgul & Suder
6. Suwirya, Candiasa & Dantes
7. Takenova & Guleva
8. Rastogi, Sharma, Mukherji & Baghel
9. Adesina, Adewuyi, Morenikeji, Ogundele & Babatunde

بررسی شده است. یکی از مطالعات (عبدالطیف، سماراسینق و ثیلاکاراتنه[1]، ۲۰۲۳) براساس ریسک مناطق، پروسه‌های امنیتی بیشتر را پیشنهاد کرده است. مطالعه‌ای دیگر (روکپاکاونگ، سابسمبون و نیلپانیچ[2]، ۲۰۲۲) نیز به افزایش امنیت با استفاده از یک مدل احراز هویت مکان محور پرداخته است. مدیریت میزان بارگذاری پول نقد در هر خودپرداز با توجه به محل قرارگیری آن، موضوع بسیاری از مطالعات بوده است (ترنگ، سنب و گیانج[3]، ۲۰۱۹)؛ زیرا میزان نشست پول در دستگاه باعث تحمیل هزینه به بانک شده و از سوی دیگر، نبود پول سبب تحمیل هزینه جریمه ناشی از تراکنش‌های ناموفق می‌شود. همچنین در مقاله مواتیسیکا[4] (۲۰۱۶)، رضایت مشتریان بانکی از خدمات بانک با استفاده از موضع خودپردازها بررسی شده است. در مطالعه دیگری، ارتباط میزان باکتری با محل استقرار خودپردازها بررسی شده است (یوانا و همکاران[5]، ۲۰۲۲). با فرض در دست داشتن چنین اطلاعاتی از بانکی دیگر، می‌توان میزان باکتری‌های موجود در خودپردازها را تخمین زد. مهم‌تر از همه، امکان تشخیص تراکنش‌های سرقت با در دست داشتن مکان خودپردازها میسر است (جاگ و پردشی[6]، ۲۰۱۴) و با توجه به نوع الگوریتم‌ها می‌توان چنین نتیجه گرفت که هر چه تعداد دستگاه‌هایی که یک بانک توانسته است شناسایی کند بیشتر باشد، احتمال موفقیت الگوریتم‌های جلوگیری از سرقت کارت افزایش می‌یابد.

در نتیجه، هر یک از بانک‌ها، به‌طور مستقل، پایگاه دادهٔ خودپردازهای متعلق به خودشان را دارند؛ اما نبود سیستم یکپارچه، باعث عدم دسترسی به مکان جغرافیایی تمامی خودپردازهای موجود در یک محل یا یک شهر شده است. از این رو نمی‌توان تحلیل‌هایی از جنس سرقت، پیش‌بینی محل مناسب برای قرار دادن دستگاه بعدی و غیره را انجام داد. یکی از راه‌حل‌های این مشکل، ایجاد پایگاه داده یکپارچه و منسجم بین تمامی بانک‌ها و مؤسسه‌های مالی است. این رویکرد بسیار اساسی و دقیقی است؛ اما با توجه به نگاه موجود که این اطلاعات را محرمانه می‌داند و تمایلی برای همکاری وجود ندارد، در کوتاه‌مدت عملی به‌نظر نمی‌رسد. در این میان می‌توان از راه‌های کوتاه‌مدت با دقت کمتر استفاده کرد. پیش‌بینی مکان خودپردازهای دیگر بانک‌ها و مؤسسه‌های مالی با استفاده از موقعیت جغرافیایی دستگاه‌های معلوم، یکی از این روش‌هاست که هدف اصلی پژوهش حاضر است. به بیان بهتر، هدف این پژوهش یافتن روشی عملی است که بتوان با استفاده از داده‌های در دسترس، مکان خودپردازهای دیگر را شناسایی کرد. سؤالی که در این پژوهش به آن پاسخ داده می‌شود، این است که آیا با استفاده از تلفیق الگوریتم‌های مبتنی بر مسیر و روش‌های پیش‌بینی یال (که فقط بر تراکنش‌های بانکی استوار است) می‌توان روشی ابداع کرد که به‌صورت کارا و با دقت کافی مکان خودپردازهای دیگر را تخمین زند.

با توجه به دانش فعلی نویسندگان، این پژوهش اولین مطالعه‌ای است که در آن از روش‌های مبتنی بر گراف، برای تخمین مکان خودپردازها استفاده شده است. با توجه به این نکته که در کشورهای دیگر، وجود پایگاه داده‌های جامع،

1. Alabdulatif, Samarasinghe & Thilakarathne
2. Rukpakavong, Subsomboon & Nilpanich
3. Trang, Sonb & Giangc
4. Mwatsika
5. Yuana et al.
6. Jog & Pardeshi

آنان را از حل چنین مسئله‌ای بی‌نیاز می‌کند، مطالعه‌ای مشابه در میان مقاله‌های خارجی یافت نشد. در کشورمان نیز این اولین بار است که چنین مسئله‌ای در سیستم‌های بانکداری کشور مطرح شده است. در ادامه، ابتدا ادبیات موضوع مرور می‌شود، سپس روش پیاده‌سازی و اعتبارسنجی تبیین شده است و بعد از ذکر یافته‌های پژوهش، به جمع‌بندی پرداخته خواهد شد.

پیشینهٔ پژوهش

گراف یکی از ابزارهای قابل اتکا در تحلیل بسیاری از مسائل از جمله فیزیک، شیمی، جامعه‌شناسی و علوم کامپیوتری است (عالم‌تبریز، طلایی و مرادی، ۱۳۹۲). در رویکرد گرافی، نگاشت مسئله به گراف از طریق نگاشت مؤلفه‌های سیستم به گره‌ها و روابط به یال‌ها صورت می‌پذیرد (سنگبر، صافی، آذر و ربیعه، ۱۴۰۰). پیش‌بینی یال در گراف، یکی از مسائل جامع در علوم کامپیوتر است که به سؤال‌های مهمی در حوزه‌های متنوع پاسخ داده است (برهمند، نصیری، رستمی و فروزنده[1]، ۲۰۲۱؛ هوانگ، لی و چن[2]، ۲۰۰۵). با توجه به پویایی شبکه، می‌توان الگوریتم‌ها را در دو دسته طبقه‌بندی کرد. در شبکه‌های ایستا، هدف، یافتن یال‌هایی است که گزارش نشده‌اند. برای مثال، در حوزهٔ علوم زیستی، بررسی وجود هر یک یال در شبکه‌های تراکنش پروتئین ـ پروتئین در محیط آزمایشگاه، بسیار پُرهزینه و زمان‌بَر است، اما با استفاده از تراکنش‌های کشف شده موجود در محیط آزمایشگاهی و تشکیل شبکه، می‌توان بقیه تراکنش‌ها را با الگوریتم‌های پیش‌بینی یال به‌دست آورد (لی و روان[3]، ۲۰۱۳). در شبکه‌های پویا، هدف، پیش‌بینی یال‌ها در طول زمان است، برای مثال از شبکه‌های اجتماعی، در شبکه ایمیل‌های یک سازمان ممکن است دو فردی که تاکنون به یکدیگر ایمیل نداده‌اند، در آینده‌ای نزدیک به یکدیگر ایمیل دهند یا دو فردی که تابه‌حال یک توییت را بازتوییت نکرده‌اند، در آینده‌ای نزدیک، به‌طور مشترک به بازتوییت یک پیام اقدام کنند. این الگوریتم‌ها می‌توانند در تشخیص جوامعی که در آینده نزدیک احتمال تشکیل دارند کمک کننده باشند (صفدری، کونتیسیانی و دیباکو[4]، ۲۰۲۲).

تکنیک‌های بسیاری در زمینه پیش‌بینی یال پیشنهاد شده است که می‌توان آن‌ها را در سه دسته اصلی تخمین بیشینه درست‌نمایی، روش‌های احتمالاتی و روش‌های مبتنی بر شباهت طبقه‌بندی کرد. در روش تخمین بیشینه درست‌نمایی، فرض می‌شود ساختار اصلی گراف از قاعده و پارامترها خاصی پیروی می‌کند که قوانین و پارامترهای حاکم بر آن با استفاده از بیشینه کردن احتمال مشاهده ساختار موجود محاسبه می‌شوند. سپس احتمال یال‌هایی که وجود ندارند با استفاده از قواعد به‌دست‌آمده محاسبه می‌شود (کلاوزت، مور و نیومن[5]، ۲۰۰۸). یکی از مشکلات عمدهٔ این روش‌ها زمان اجرای طولانی آن‌هاست. این روش‌ها در شبکه‌هایی با حدود چند هزار گره، ممکن است در زمان قابل قبولی پاسخ دهند؛ اما در شبکه‌های برخط که ممکن است چند میلیون گره داشته باشند، زمان اجرا بسیار طولانی خواهد شد. علاوه‌براین پاسخ‌هایی که از این روش به دست می‌آیند نیز چندان دقیق نیست (لو و ژو[6]، ۲۰۱۱). در مدل‌های

1. Berahmand, Nasiri, Rostami & Forouzandeh
2. Huang & Chen
3. Lei & Ruan
3. Safdari, Contisciani & De Bacco
5. Clauset, Moore & Newman
6. Lü & Zhou

احتمالاتی، ابتدا ساختار موجود در شبکه مطالعه و آموخته می‌شود؛ سپس احتمال یال‌ها با مدل آموخته‌شده محاسبه می‌شوند. برای مثال در مدل رابطه‌ای احتمالی (PRM)[1]، توزیع احتمال توام بین ویژگی‌های داده‌های مرتبط محاسبه می‌شود (هکرمن، میک و کولر[2]، ۲۰۰۴). این روش در واقع نسخه رابطه‌ای توسعه‌یافته از شبکه‌های بیزی است که هرچند در برخی از حوزه‌ها به‌خوبی عمل می‌کند، در موارد دیگر که دورهای زیادی در گراف وجود داشته باشد، به تشکیل دور در شبکه بیزی منجر می‌شود که دیگر نمی‌توان از آن مدل منسجم احتمالاتی به حساب آورد (تسکار، ابیل، ونگ و کولر[3]، ۲۰۰۷). دسته دیگری که الگوریتم‌های زیادی را شامل می‌شود روش‌های مبتنی بر شباهت است که در آن‌ها شباهت بین هر جفت گره بررسی شده و از این طریق احتمال حضور هر یال محاسبه می‌شود. بر خلاف ایده به ظاهر ساده این دسته از روش‌ها، تعریف شباهت بین هر جفت گره موضوع مهم و چالش‌برانگیزی است تا آنجا که بسته به تعریف معیار شباهت، ممکن است این روش‌ها سودمند یا به‌طور کامل بی‌فایده باشند. در این مطالعه با تعریف معیار شباهت، یک فیلتر برای حذف یال‌های پیش‌بینی شده ایجاد شد.

حال همان گونه که پیش‌تر ذکر شد با استفاده از اطلاعات مکانی و زمانی موجود می‌توان به پیش‌بینی یال‌های جدید پرداخت. این ویژگی‌های مکانی و زمانی با استفاده از الگوریتم‌های مبتنی بر مسیر به‌دست می‌آیند. لذا در این مطالعه با استفاده از موقعیت مکانی دستگاه‌های خودپرداز که در بانک اطلاعاتی موجود بود و نیز فاصله زمانی بین تراکنش‌های میان دو دستگاه، تلاش شد تا موقعیت مکانی دستگاه‌های جدید شناسایی شود. از آنجا که نام بانک و شمارهٔ خودپرداز در تراکنش‌ها موجود است و از سوی دیگر، با استفاده از تراکنش‌های یک کارت مشخص میان دو دستگاه، فاصلهٔ زمانی دو دستگاه نیز قابل محاسبه است، از ترکیب این دو ویژگی برای شناسایی موقعیت قرارگرفتن دستگاه‌ها و برچسب زدن به آن‌ها استفاده شد. داده‌کاوی زمانی ـ مکانی کلیدی‌ترین عنصر تمامی این روش‌هاست (پائو، فدلیل، لین و چن[4]، ۲۰۱۲؛ یوان، ژنگ، ژی و سان[5]، ۲۰۱۱). برای مثال در مطالعاتی، جابه‌جایی موقعیت مکانی کارت شناسایی دانشجویان بررسی شده و از روی آن محل بعدی که حضور خواهند یافت، پیش‌بینی شده است (باراباسی[6]، ۲۰۰۵؛ بروکمن، هوفنیگل و گیسل[7]، ۲۰۰۶).

الگوریتم‌های پیش‌بینی یال از کارایی خوبی برخوردارند؛ اما برای حل موضوع مورد بررسی این مطالعه، یعنی پیش‌بینی مکان خودپردازها، به‌تنهایی کافی نیستند؛ زیرا در الگوریتم‌های مذکور، وجود یا عدم وجود یال، نشانهٔ میزان فاصلهٔ جغرافیایی نیست و گاهی فقط با استفاده از این اطلاعات مکانی، دقت پیش‌بینی یال افزایش یافته است. بنابراین در گام بعدی، الگوریتم‌های مبتنی بر مسیر مرور خواهد شد.

در سال‌های اخیر، با توسعه نرم‌افزارهای شبکهٔ اجتماعی که در آن، افراد موقعیت مکانی خود را با دیگران به

1. Probabilistic relational models
2. Heckerman, Meek & Koller
3. Taskar, Abbeel, Wong & Koller
4. Pao, Fadlil, Lin & Chen
5. Yuan, Zheng, Xie & Sun
6. Barabasi
7. Brockmann, Hufnagel & Geisel

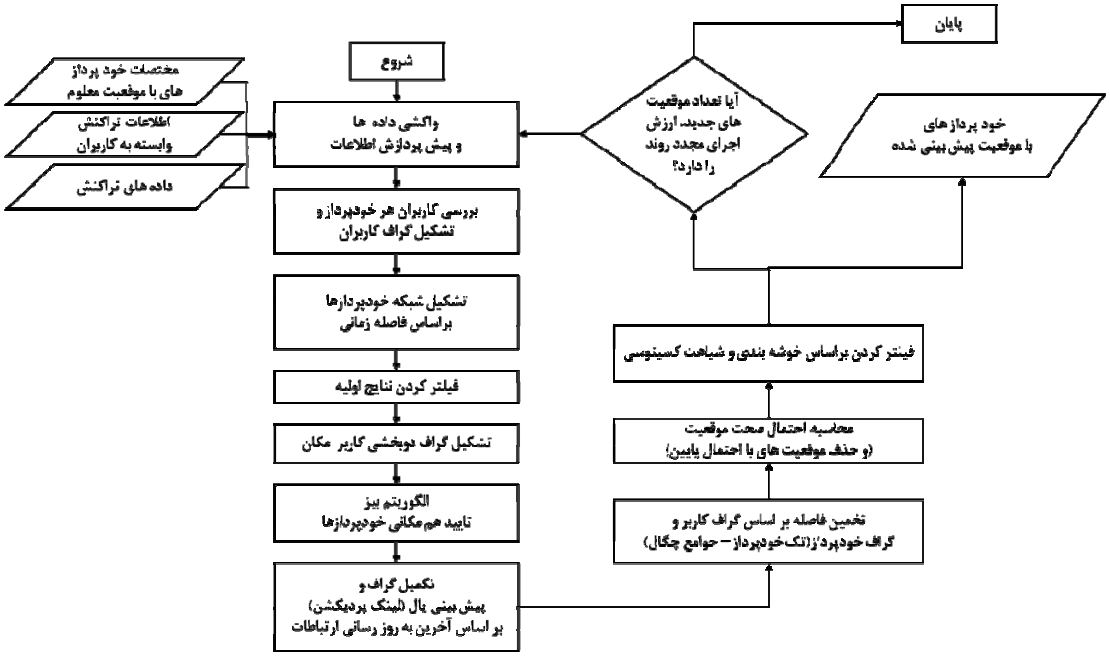
اشتراک می‌گذارند، پیش‌بینی یال در شبکه‌های اجتماعی مبتنی بر مکان (LBSN)[۱]، بسیار در کانون توجه قرار گرفته است (نیومن[۲]، ۲۰۰۱؛ آدامیک و آدار[۳]، ۲۰۰۳؛ لیبن نوول و کلاینبرگ[۴]، ۲۰۰۳؛ ژنگ، ژنگ، ما، زی و ما[۵]، ۲۰۱۱). برای پیش‌بینی یال در این شبکه‌ها، برخی از دانشمندان بدون استفاده از ویژگی‌های مکانی، روی روش‌های مرسوم در این زمینه تمرکز کرده‌اند؛ مانند تأثیر وزن‌دار تعداد همسایگان مشابه (نصیری، برهمند، سامعی و همکاران[۶]، ۲۰۲۲)؛ در حالی که عده‌ای دیگر، اطلاعات مکانی‌ای را که به اشتراک گذاشته شده نیز در نظر گرفته‌اند (وی و همکاران[۷]، ۲۰۲۳). در مطالعه‌ای که در دانشگاه کمبریج انجام شد، به این نتیجه دست یافتند که استفاده از ویژگی‌های ساختاری همچون تعداد همسایگان مشابه، نسبت درجه ورودی به خروجی یا همسایگی با درجه دو، به‌تنهایی کافی نیست، بلکه استفاده از اطلاعات مکانی افرادی که هم‌زمان در یک مکان شروع به اشتراک اطلاعات کرده‌اند، دقت نتایج را به صورت قابل ملاحظه‌ای افزایش می‌دهد (آلامانیس، سلاتو و ماسکولو[۸]، ۲۰۱۲). در مطالعه‌ای دیگر در حوزه این شبکه‌ها، پس از مروری مفصل بر الگوریتم‌های متعدد پیش‌بینی لینک، بر اهمیت استفاده از اطلاعات پایهٔ مکانی، به‌عنوان یک ویژگی مهم در پیشنهاد درست، تأکید شده و غنی‌سازی داده‌ها با اطلاعات مکانی به بهبود شناسایی‌ها منتج شده است (ژانگ، لی، سونگ، دافیلد و ژانگ[۹]، ۲۰۲۳). شایان ذکر است نکته مهمی که در تمامی مطالعات مشاهده می‌شود که استفاده از داده‌های مکان ـ زمان برای افزایش دقت پیش‌بینی است. منظور از داده‌های مکان ـ زمان، داده‌هایی است که مکان هر کاربر را در زمان معینی نشان می‌دهد. در مطالعه پیش رو با استفاده از این روش‌ها، شبکه خودپردازها به شبکه‌ای مبتنی بر مکان تبدیل شد.

روش‌شناسی پژوهش

این مطالعه از نوع اکتشافی ـ پیمایشی است و پژوهشی کاربردی محسوب می‌شود. طبق مطالعه پیشین (نرایانان، قروکوری[۱۰]، ۲۰۱۶) برای تشکیل شبکه‌های LSBN سه نوع داده نیاز است: اطلاعات کاربران، تاریخچه فعالیت و مکان جغرافیایی فعالیت آن‌ها. در این پژوهش از سه نوع دادهٔ کاربران، دادهٔ تراکنش‌ها و اطلاعات بخشی از خودپردازهای موجود در پایگاه دادهٔ شرکت داتیس آرین قشم (داتین) استفاده شد. با استفاده از اطلاعات کاربران، شناسه مشتری و شمارهٔ کارت‌های مرتبط با هر مشتری استخراج شد. طول و عرض جغرافیایی، نام شهر و استان محل استقرار ۲۱۰۰ خودپرداز که در سراسر ایران پراکنده است، از پایگاه داده مرتبط با خودپردازها به دست آمدند. در لایهٔ بعدی، ارتباط بین مشتریان و دستگاه‌های خودپرداز با استفاده از تراکنش‌ها به‌دست آمد. به این منظور ۵۸۹۳۰۳۹۹۸ تراکنش ثبت شده در طول چهار ماه بررسی شد.

1. Location-Based Social Network
2. Newman
3. Adamic & Adar
4. Liben-Nowell & Kleinberg
5. Zheng, Zhang, Ma, Xie & Ma
6. Nasiri, Berahmand, Samei & Li
7. Wei et al.
8. Allamanis, Scellato, & Mascolo
9. Zhang, Li, Song, Duffield & Zhang
10. Narayanan & Cherukuri

برای بررسی صحت نتایج، داده‌ها به دو بخش مجموعه داده آموزش و مجموعه داده راستی‌آزمایی تقسیم شدند. بنابراین مجموعه داده آموزش شامل ۱۶۸۰ خودپرداز و مجموعه داده راستی‌آزمایی شامل ۴۲۰ خودپرداز است. بنابراین در نهایت دقت الگوریتم با بررسی نتایج روی ۴۲۰ دستگاه خودپرداز که از ابتدا کنار گذاشته‌ایم سنجیده می‌شود.



شکل ۱. فلوچارت تخمین مکان جغرافیایی

تشکیل گراف کاربر ــ مکان

شکل ۲ مراحل کلی اجرای این پژوهش است. همان گونه که پیش‌تر ذکر شد، اولین گام تشکیل گراف کاربر ـ مکان از نوع LBSN است (شکل ۱– الف). برای تشکیل شبکه کاربران $(G_u = (U, E_u))$، کاربران با نماد U، به‌عنوان گره در نظر گرفته شده است و یال‌ها با نماد E، از طریق فرمول زیر به دست آمده است:

$$E_u \subseteq \{x, y \mid x, y \in U \; and \; x \neq y \; and \; ATM_{x=} \; ATM_y\}$$ رابطهٔ ۱)

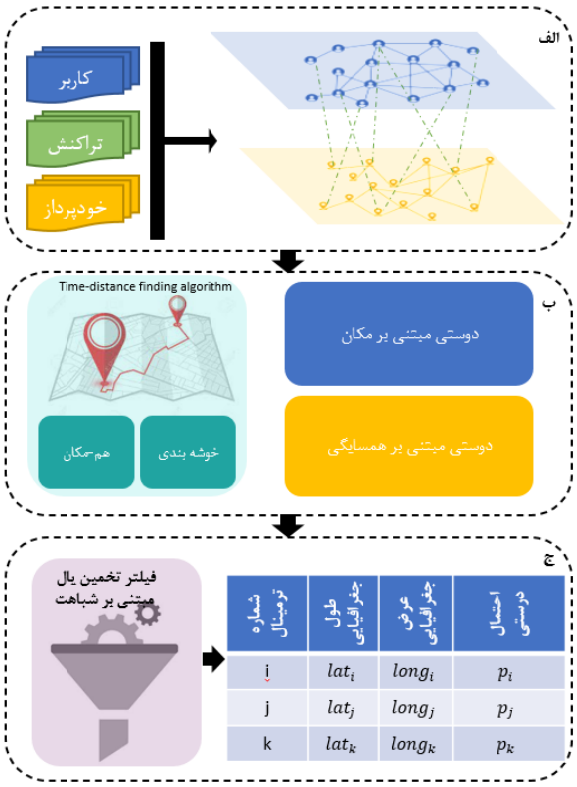
ATM_i، دستگاه خودپردازی است که کاربر i از آن استفاده کرده است. طبق تعریف، اگر هر دو کاربر از یک دستگاه خودپرداز برای تراکنش‌ها استفاده کرده باشند، به‌عنوان دوست در شبکه اجتماعی در نظر گرفته شدند.

در مرحله بعدی، شبکه خودپردازها $(G_{ATM} = (ATM, E_{ATM}))$ با استفاده از فاصله بین آن‌ها $(D_{x,y})$ تشکیل شد.

$$D_{x,y} = min_{u \in U}(\sqrt[2]{(t(TRX_u^x) - t(TRX_u^y))^2})$$ رابطهٔ ۲)

$$E_{ATM} \subseteq \{x, y \mid x, y \in ATM \; and \; x \neq y \; and \; D_{x,y} \leq C\}$$ رابطهٔ ۳)

فاصله محاسبه شده از جنس زمان (t) است. در رابطهٔ ۲، TRX_i^j تراکنشی است که کاربر i روی خودپرداز j انجام داده و $t(TRX_j^i)$ زمان انجام تراکنش کاربر i روی خودپرداز j است. در محاسبه مینیمم مقدار زمانی، اگر فاصله مینیمم از بقیه فواصل زمانی بیشتر از یک چارک بود، دومین مینیمم به‌عنوان فاصله زمانی در نظر گرفته شد. فواصل زمانی بسیار بزرگ نیز حذف شدند.



شکل ۲. مراحل کلی اجرای پژوهش.

الف) ابتدا گراف دو بخشی کاربر ـ مکان تشکیل شد. ب) با استفاده از این دو گراف شباهت مکانی و شباهت همسایگی محاسبه شد. در همین مرحله، الگوریتم یافتن فاصله زمانی ـ مکانی که خود شامل دو مرحله است، اجرا شد و فاصله به یال‌ها تخصیص یافت. ج) یال‌های به‌دست‌آمده فیلتر شده و برای هر خودپرداز مختصات جغرافیایی و احتمال درستی گزارش شد.

با تشکیل گراف دو بخشی بین خودپردازها و کاربران $\left(G_{U,ATM} = (U, ATM, E_{u,ATM})\right)$، شبکه اجتماعی مبتنی بر مکان تشکیل شد. در این تعریف $G_{U,ATM}$ گراف دو بخشی مشتمل بر گره‌هایی از جنس خودپرداز و کاربران (ATM و U) است و یال‌های این گراف $(E_{u,ATM})$ علاوه‌بر یال‌های موجود در هر گراف کاربر و گراف خودپردازها، یال‌هایی میان کاربران و خودپردازها را نیز شامل می‌شود.

رابطهٔ ۴)
$$E_{u,ATM} \subseteq \{x, y \mid x \in U\ , y \in ATM\ and\ TRX_y^x\ \in TRX\}$$

الگوریتم محاسبه فاصله مکان ــ زمان

گام اول در این الگوریتم، یادگیری طول و عرض جغرافیایی با استفاده از گراف خودپردازهاست. با استفاده از دسته‌بند بیز ساده[1]، فواصل زمانی که نشان‌دهندهٔ هم‌مکان بودن دو دستگاه خودپرداز است، محاسبه شد و برچسب هم‌مکان و غیرهم‌مکان، بر اساس نگاشت طول و عرض جغرافیایی روی گراف به‌دست آمد و ترکیب‌های دوتایی دستگاه‌های خودپرداز، همراه فواصل زمانی، به‌عنوان ورودی و برچسب به الگوریتم داده شد. همچنین به‌منظور پرهیز از خطاهای ایجاد شده توسط نمونه‌برداری[2]، از روش اعتبارسنجی متقابل ۵ برابر[3] استفاده شد. گام دوم، محاسبه فواصل زمانی بین دستگاه‌های معلوم و مجهول و کامل کردن گراف مکانی است. G_{ATM}، G_u و $G_{u,ATM}$ نیز مانند مرحله قبل کامل شدند.

در گام سوم، دستگاه‌ها با استفاده از قواعد به‌دست‌آمده، به دو دسته هم‌مکان و فاصله‌دار تقسیم شدند. طول و عرض جغرافیایی خودپردازهایی که در دسته هم‌مکان با خودپردازهای معلوم قرار گرفتند در همین مرحله مشخص شد. برای به‌دست آوردن موقعیت دستگاه‌های دسته دوم، با استفاده از روش تشخیص جامعه[4] در گراف ، خودپردازهایی که در یک زیرمجموعه چگال در کنار یکدیگر بودند، شناسایی شدند. خوشه‌های به‌دست‌آمده، بر اساس تعداد خودپردازهای با موقعیت معلوم به دو نوع تقسیم شدند:

۱. خوشه‌هایی بدون خودپرداز با موقعیت معلوم. برای این خوشه‌ها فقط می‌توان بر اساس نزدیک‌ترین خوشه دارای موقعیت، مختصاتی با احتمال بسیار پایین تولید کرد. بنابراین، این دسته به‌عنوان بخشی که الگوریتم توانایی تشخیص موقعیت آن‌ها را ندارد، کنار گذاشته شد.

۲. خوشه‌هایی با یک یا بیش از یک خودپرداز با موقعیت معلوم. در این نمونه مرکز خوشه برابر با میانگین جغرافیایی مختصات دستگاه‌های معلوم است. موقعیت هر نقطه نیز با استفاده از میانگین وزن‌دار به نسبت نقاط دارای مختصات تعیین می‌شود.

$$latitude_x = \sum_{known \in ATM} D_{x,known} \times latitude_{known}$$ رابطهٔ ۵)

$$longitude_x = \sum_{known \in ATM} D_{x,known} \times longitude_{known}$$ رابطهٔ ۶)

در روابط ۵ و ۶، $latitude$ و $longitude$، طول و عرض جغرافیایی هر دستگاه و D فاصله زمانی هر دستگاه با دستگاه‌های معلوم است.

محاسبه احتمال صحت موقعیت‌های به‌دست‌آمده

بعد از انتصاب موقعیت به هر یک از دستگاه‌ها، لازم است احتمال صحت آن‌ها بررسی شود. در LSBN ها با استفاده از

1. Naive Bayes classifier
2. Sampling
3. 5-fold Cross-Validation
4. Community Detection

مفهوم دوستی در شبکه، احتمال اینکه هر فرد به کدام موقعیت بعدی مهاجرت می‌کند، محاسبه می‌شود. در مطالعهٔ لی، گی، هونگ و ژو[1] (۲۰۱۶) مشاهده شد افرادی که در نزدیکی یکدیگر زندگی می‌کنند، با احتمال بالاتری در محل‌های یکسان حضور دارند. علاوه‌بر تشکیل شبکه کاربران بر این اساس، احتمال دوستی همسایگی[2] بین دو دستگاه x و y به‌صورت زیر تعریف شد:

$$NF_{x,y} = \frac{\sum_{u_i \in TRX_x^i \; AND \; u_i \in TRX_y^i} U_i}{\sum_{u_i \in TRX_x^i \; OR \; u_i \in TRX_y^i} U_i}$$ رابطهٔ ۷)

طبق این تعریف تعداد کاربرانی که در گراف کاربر با یکدیگر دوست هستند و از طریق این دو دستگاه تراکنشی انجام داده‌اند، نسبت به جمع کاربرانی که روی هر یک تراکنشی انجام داده‌اند، به‌عنوان احتمال دوستی همسایگی در نظر گرفته می‌شود.

از طرف دیگر احتمال اینکه کاربرانی که در محل‌های یکسان حضور پیدا کنند، همچنان اماکن یکسانی را انتخاب کنند، بیشتر است. بنابراین پارامتر دیگری به نام دوستی موقعیتی نیز تعریف شد. به عبارت بهتر، ابتدا با استفاده از تراکنش‌ها، مکان‌های پرتردد هر کاربر پیدا شد؛ سپس با استفاده از آن احتمال نزدیک بودن دو دستگاه بر اساس دوستی موقعیتی کاربرانی که از آن‌ها استفاده کرده‌اند، به دست آمد. ابتدا ترددهای هر کاربر در مدت زمان مشخصی بر اساس ساعات شبانه روز مورد بررسی قرار گرفت. به ازای هر خودپرداز، تعداد دفعات استفاده توسط کاربر مشخص شد و سپس خودپردازهایی که به لحاظ تعداد در چارک سوم و چهارم قرار داشتند، به‌عنوان خودپردازهای کاندید به‌دست آمدند. در مرحله بعد شباهت میان کاربران بر اساس خودپردازهای بیشتر استفاده شده توسط آن‌ها و تعداد استفاده از هر خودپرداز اندازه‌گیری شد. با استفاده از این معیار شباهت، شبکه کاربران دوباره تشکیل شد و معیار دوستی موقعیتی[3] نیز همانند دوستی همسایگی محاسبه شد.

$$LF_{x,y} = \frac{\sum_{u_i \in TRX_x^i \; AND \; u_i \in TRX_y^i} U_i}{\sum_{u_i \in TRX_x^i \; OR \; u_i \in TRX_y^i} U_i}$$ رابطهٔ ۸)

در نهایت با استفاده از میانگین دو معیار NF و LF، میزان احتمالی به هر یک از یال‌ها اختصاص یافت.

فیلتر بر اساس معیار شباهت

روش‌های پیش‌بینی یال، روش‌هایی هستند که با استفاده از آن‌ها به گراف یال‌هایی اضافه و یا بخشی از یال‌ها حذف می‌شوند. اخیراً استفاده از این روش‌ها به‌عنوان فیلتر در ادغام با سایر روش‌ها، به نتایج با دقت بالاتری منجر شده است (هوانگ و همکاران، ۲۰۰۵). از این رو در این بخش با استفاده از روش پیش‌بینی یال بر اساس شباهت گره‌ها، یال‌های شبکه فیلتر شدند.

الگوی ساعتی استفاده از خودپردازها در هر محله و هر شهر با محله و شهر دیگر متفاوت است. برای مثال،

1. Li, Ge, Hong & Zhu
2. Neighbouring Friendship
3. Location Friendship

خودپردازهای واقع در محله ولی‌عصر تهران، حتی در ساعات پایانی روز همچنان در حال خدمات‌دهی هستند. این در حالی است که احتمال استفاده از خودپردازهای واقع در شهرهای کوچک‌تر در همان ساعت بسیار پایین است. بر این اساس، میانگین تعداد تراکنش‌های هر خودپرداز در هر روز محاسبه و خودپردازها با استفاده از خوشه‌بندی سلسله‌مراتبی دسته‌بندی شدند؛ سپس یک بردار ۲۴ تایی گسسته برای هر خودپرداز با استفاده از میانگین تعداد تراکنش‌ها در هر ساعت که نسبت به تعداد کل تراکنش‌ها نرمال شده بودند، به دست آمد. با استفاده از شباهت کسینوسی، میزان شباهت رفتاری خودپردازها محاسبه شد. بنابراین با استفاده از دو فیلتر دندوگرام و سپس شباهت کسینوسی کمتر از حد آستانه، یال‌ها فیلتر شدند. در نهایت میزان احتمال روی هر یال، به‌عنوان میزان صحت موقعیت جغرافیایی پیش‌بینی شده گره متصل به آن گزارش شد.

اعتبارسنجی

همان‌گونه که پیش‌تر ذکر شد، خروجی الگوریتم ATMLP به صورت Loc(ATM)=(lat,long,pr) است. برای هر یک از ATMها فاصله طول و عرض جغرافیایی پیش‌بینی شده با طول و عرض جغرافیایی حقیقی از طریق رابطه زیر محاسبه شد.

$$D(ATM) = cos^{-1}(sin\, lat_{actual} \times sin\, lat_{predicted} + cos\, lat_{actual} \times cos\, lat_{actual} \times cos\, long_{predicted} - long_{actual}) \times R$$

رابطهٔ ۹)

$$validity = \frac{\sum_{D(ATM_i) \le 50} p(ATM_i)}{\sum_i ATM_i}$$

رابطهٔ ۱۰)

اعتبار الگوریتم بر اساس جمع احتمال تمامی دستگاه‌هایی که فاصله پیش‌بینی شده از مقدار واقعی کوچک‌تر از ۵۰ متر است محاسبه شد. نرخ خطا نیز از طریق رابطهٔ زیر محاسبه شد.

$$Error = \frac{\sum_{D(ATM_i) \le 50}(1 - p(ATM_i)) + \sum_{D(ATM_i) > 50} ATM_i}{\sum_i ATM_i}$$

رابطهٔ ۱۱)

یافته‌های پژوهش

به‌منظور یافتن مکان جغرافیایی دستگاه‌های خودپرداز، گراف خودپردازها و گراف کاربر تشکیل شد. منظور از گراف کاربران، گرافی است که بر اساس تراکنش‌های موجود و محل تراکنش، کاربران را به یکدیگر مرتبط می‌کرد و منظور از گراف خودپردازها، تشکیل گرافی با استفاده از فاصله است. با برقراری ارتباط بین این دو گراف، شبکه کاربر ـ مکان به دست آمد. همان گونه که در بخش پیشین توضیح داده شد، پنج طبقه‌بند بیز ساده آموزش داده شد. ورودی این پنج طبقه‌بند که مشتمل بر فاصله زمانی دو دستگاه، فاصله جغرافیایی دو دستگاه و تعداد دوست‌های اجتماعی مشترک است، از شبکه کاربر ـ مکان استخراج شد. با استفاده از فاصله جغرافیایی، به هر جفت دستگاه برچسب هم مکان و فاصله‌دار منتصب شد. برای مثال جدول ۱ را در نظر بگیرید. با اینکه فاصله زمانی ثبت شده بین دو دستگاه در رکورد شماره ۲

کوتاه‌تر از جفت دستگاه رکورد شماره ۱ است؛ اما در واقع از یکدیگر دور هستند و تنها دو نفر بین این دو دستگاه تراکنش مشترکی انجام داده‌اند. دقت شود که برای سنجش میزان اعتبار الگوریتم از ۲۱۰۰ دستگاه خودپردازی که مکان جغرافیایی معلوم داشتند، ۴۲۰ عدد به‌عنوان داده‌های راستی آزمایی کنار گذاشته شد. بنابراین تا قبل از مرحله راستی‌آزمایی تنها می‌توان از ۱۶۸۰ دستگاه باقی‌مانده استفاده کرد. برای آموزش دسته‌بند، از روش اعتبار سنجی متقابل ۵ برابر استفاده شد و در نهایت بهترین دسته‌بند برای مراحل بعدی انتخاب شد. دستگاه‌هایی با فاصله بیشتر از ۱۰۰ متر، به‌عنوان فاصله‌دار در نظر گرفته شدند. تعداد ۱۳۹۹× ۱۴۰۰ نمونه برای آموزش دسته‌بند استفاده شد. از آنجا که تعداد نمونه‌های فاصله‌دار بسیار بیشتر از نمونه‌های هم‌مکان بود، برای برقراری توازن بین تعداد نمونه‌های هر دسته، نمونه‌های فاصله‌دار به‌صورت تصادفی به تعداد تمامی نمونه‌های هم مکان انتخاب شدند.

جدول ۱. نمونه‌ای از برچسب‌های هم‌مکان و فاصله‌دار

ورودی			خروجی	
شماره جفت دستگاه	تعداد دوست‌های اجتماعی مشترک	فاصله زمانی (ثانیه)	فاصله جغرافیایی (متر)	برچسب
۱	۲۰	۳۵۰	۹۰	هم مکان
۲	۲	۳۰۰	۱۵۰	فاصله‌دار

بعد از آموزش دسته‌بند، داده‌های آزمایش (۲۸۰ خودپرداز) به‌عنوان ورودی به این دسته‌بندها داده شد. نتایج آموزش پنج نوبت آموزش طبقه‌بند بیز در جدول ۲ نمایش داده شده است.

جدول ۲. نتایج آموزش پنج نوبت آموزش طبقه‌بند بیز

نوبت	دقت	صحت	بازیابی	حساسیت
۱	۹۵/۴	۹۳/۷	۹۷/۱	۹۳/۵
۲	۹۳/۹	۹۱/۸	۹۶/۴	۹۱/۴
۳	۹۶/۴	۹۴/۵	۹۸/۵	۹۴/۲
۴	۹۵/۷	۹۳/۸	۹۷/۸	۹۳/۵
۵	۹۶/۴	۹۳/۹	۹۹/۲	۹۳/۵
میانگین	۹۵/۶	۹۳/۶	۹۷/۹	۹۳/۳

بعد از اطمینان از دقت و صحت بالای دسته‌بند، ۴۲۰ خودپردازی که به‌عنوان دادهٔ راستی‌آزمایی کنار گذاشته شده بود، به شبکه اضافه شد. سپس تمامی ترکیب‌های دوتایی این ۴۲۰ خودپرداز با بقیه خودپردازها (دستگاه‌هایی با مکان جغرافیایی معلوم) تشکیل شد. ترکیب‌هایی که در شبکه اصلاً یالی بین آن‌ها وجود نداشت، حذف شدند و بقیه به دسته‌بند بیزی که آموزش داده شده بود، وارد شدند. ۲۰۳ دستگاه با طول و عرض جغرافیایی مجهول تنها با یک دستگاه با طول و عرض جغرافیایی معلوم هم مکان تشخیص داده شدند. بنابراین همان مکان دستگاه‌های معلوم به مجهول منتصب شد. بقیه دستگاه‌ها، یا با بیش از یک دستگاه معلوم هم مکان تشخیص داده شده بودند یا با تمامی دستگاه‌های معلوم فاصله‌دار تشخیص داده شده بودند. برای حل این دو مشکل، زیرمجموعه‌های چگال، در گراف به‌دست آمدند (جدول۳).

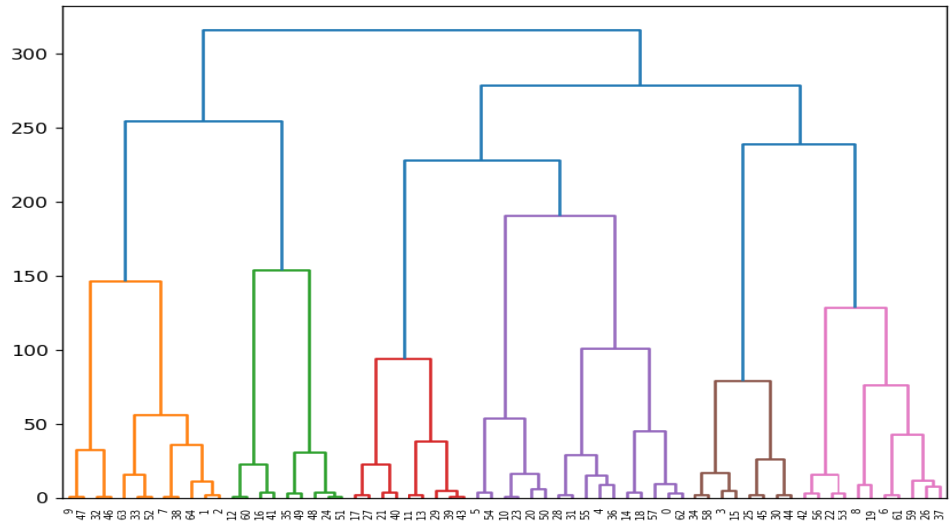
جدول ۳. زیرمجموعه‌های چگال

	تعداد زیر مجموعه‌های چگال هم مکان	تعداد خودپردازها
زیرمجموعه‌های چگال که بیشتر از یک دستگاه خودپرداز با موقعیت مکانی معلوم در آن وجود دارد.	۸	۱۰۷
زیرمجموعه‌های چگال که هیچ دستگاه خودپرداز با موقعیت مکانی معلوم در آن وجود ندارد.	۱۰	۱۳

برای ۱۳ خودپرداز، امکان انتصاب طول و عرض جغرافیایی با احتمال صحت بالا وجود ندارد. هرچند به آن‌ها نیز طول و عرض نزدیک‌ترین همسایهٔ چگال منتصب شد؛ اما در نهایت به علت احتمال پایین صحت، در نتایج به‌عنوان عدم توانایی الگوریتم در نظر گرفته شدند.

بعد از محاسبه میانگین دو نوع معیار شباهت که در بخش پیشین توضیح داده شد، احتمالی به هر یال که در شبکه انتصاب شد و یال‌هایی با احتمال کوچک‌تر از ۰/۶ حذف شدند. بعد از اعمال فیلتر شباهت، در نهایت جایگاه ۳۴۰ خودپرداز با فاصله کمتر از ۱۰۰ متر با طول و عرض جغرافیایی واقعی خود و احتمال بالای ۰/۶ پیش‌بینی شد.

گام بعدی در واقع، برای بالا بردن اعتبار الگوریتم استفاده شده است. به این ترتیب که دو مرحله بیشتر برای فیلتر کردن یال‌ها به کار برده شد. ابتدا شباهت کسینوسی بین هر دو دستگاه بر اساس معیارهای ساعت و مبلغ محاسبه شد و ارتباط بین دستگاه‌هایی که شباهتی کوچک‌تر از ۰/۶ داشتند، حذف شد. علاوه‌براین، خوشه‌بندی‌ای بر اساس دو معیار ذکر شده انجام شد و باز یال میان دستگاه‌هایی که در خوشه‌های مختلف قرار داشتند، حذف شد. این دو مرحله باعث افزایش دقت الگوریتم می‌شود.



شکل ۳. خوشه‌بندی سلسله‌مراتبی میانگین تعداد تراکنش‌های هر خودپرداز در یک روز

در این مرحله اعتبار الگوریتم بر اساس معیار تعریف شده در بخش پیشین برای ۴۲۰ دستگاه خودپرداز که برای راستی آزمایی کنار گذاشته شده بود، به‌دست آمد. محل تعداد ۳۸۶ دستگاه خودپرداز، یعنی ۹۱/۹ درصد، به‌درستی

پیش‌بینی شد. اعتبار به‌دست‌آمده برابر ۸۰/۹۵ است. لازم است یادآوری شود که این عدد نشان دهندهٔ تعداد خودپردازهای درست نیست. به عبارت دیگر، اگر مکان جغرافیایی دستگاهی به‌طور درست پیش‌بینی شده باشد؛ اما احتمال پیش‌بینی برابر ۰/۷ باشد، به‌جای عدد ۱ تنها ۰/۷ به اعتبار تخصیص داده می‌شود. به همین علت معیار اعتبار تعریف شده کمتر از درصد خودپردازهایی است که به‌درستی پیش‌بینی شده‌اند. خطای این الگوریتم برابر ۱۵/۰۵ است. بنابر تعریف، خطا برابر مکمل احتمالی دستگاه‌هایی است که به‌درستی تشخیص داده شده‌اند، به‌علاوهٔ احتمال دستگاه‌هایی که به اشتباه شناسایی شده‌اند.

در گام بعدی، این الگوریتم روی خودپردازهایی که در واقع طول و عرض جغرافیایی آن‌ها در دسترس نبود، اجرا شد و با استفاده از ۲۱۰۰ خودپرداز، طول و عرض جغرافیایی ۴۰۰۰ خودپرداز مربوط به بانک‌های دیگر شناسایی شد.

نتیجه‌گیری و پیشنهادها

با توجه به اینکه محل استقرار دستگاه‌های خودپرداز، در سیستم مالی دادهٔ کلیدی محسوب می‌شود، هدف این مطالعه یافتن موقعیت جغرافیایی دستگاه‌های خودپرداز بوده است. به این منظور سه مرحله اساسی انجام شد. در گام اول، گراف مکان ـ کاربر تشکیل شد و با استفاده از داده‌های موجود در سطح گراف مکانی، یک سیستم دسته‌بند بیز ساده آموزش داده شد. میانگین دقت و صحت در پنج بار اجرا (اعتبارسنجی متقابل ۵ برابر)، با توجه به در نظر گرفتن دستهٔ هم‌مکان به‌عنوان دستهٔ مثبت، به‌ترتیب برابر با ۹۵/۶ و ۹۳/۶ بود. در گام بعدی، میزان احتمالی در بازه [۰,۱] به هر یک از یال‌ها اختصاص یافت.

در خوشه‌بندی تمامی ۲۱۰۰ دستگاه خودپرداز، ۶۴ خوشه مختلف شناسایی شد. اگر دو دستگاه در دسته‌های متفاوتی قرار گرفته بودند، یال میان آن‌ها حذف شد. علاوه‌براین، در گراف مکانی، یال‌های ما بین دو گره با شباهت کسینوسی کمتر از ۰/۶ نیز حذف شدند. سپس احتمال یال‌ها، به گره‌های متصل به آن منتصب شد. در نتیجهٔ فرایند فوق ۴۰۰۰ دستگاه در ۱۲ بانک شناسایی شد، اعتبار و خطای نتایج پیش‌بینی شده، به‌ترتیب برابر با ۸۰/۹۵ و ۱۵/۰۵ درصد بود. روش فوق روی حدود ۵۶۰ میلیون تراکنش در بازه زمانی ۴ ماهه از ابتدای فروردین ماه سال ۱۴۰۱ تا پایان تیرماه همان سال انجام گرفت. با توسعه زمانی روش در بازه زمانی دو ساله، می‌توان به طول و عرض جغرافیایی تعداد بیشتری از ۶۱۰۱۵ دستگاه خودپرداز موجود با روزانه ۱۶ میلیون و ۶۰۰هزار تراکنش در شبکهٔ بانکی دست یافت. همچنین با بررسی ناهنجاری داده‌های فوق، می‌توان به شناسایی تراکنش‌های تقلب و تخلف دست یافت؛ زیرا نشان‌دهنده بروز تراکنش‌هایی است که احتمال وقوع آن‌ها غیرممکن به نظر می‌رسد. در ادامه پیشنهاد می‌شود برای توسعه روش مذکور اصلاحات زیر در پژوهش‌های آتی انجام گیرد تا به بهبود نتایج منجر شود.

- موقعیت جغرافیایی بسیاری از شعب بانک که معمولاً دستگاه ATM دارند و برخی از خودپردازها در نقشه‌های آنلاین ثبت شده است. گرچه موقعیت‌های ثبت شده در همهٔ موارد صحیح یا به‌روز نیست، می‌توان با خزش[1] این اطلاعات از نقشه‌های برخط و پالایش آن‌ها، در صحت‌سنجی داده‌ها استفاده کرد.

1. Crawl

- با استفاده از برخی از اطلاعات جانبی مانند منطقه، ناحیه، شهر و استان محل نصب هر خودپرداز، می‌توان از دسته‌بندی منطقه‌ای آن‌ها در راستای بهبود نتایج الگوریتم بهره برد.
- با داشتن تراکنش مربوط به استفاده کارت در دستگاه‌های پوز ثابت و اطلاعات جغرافیایی آن دستگاه‌ها می‌توان موقعیت تعداد بیشتری از خودپردازهای سایر بانک‌ها را پیش‌بینی کرد.
- با استفاده از برخی از اطلاعات مشتریان (مانند آدرس محل زندگی) می‌توان گراف شبکه کاربران را پالایش کرد.

سپاسگزاری

مالکیت این مقاله متعلق به بخش تحقیق و پژوهش شرکت داتیس آرین قشم (داتین) است و به‌خاطر حمایت مالی و همکاری در اجرای پژوهش حاضر سپاسگزاری می‌شود.

منابع

آقا قلیزاده سیار، علیرضا؛ معتدل، محمدرضا و پورابراهیمی، علیرضا (۱۳۹۵). ارائه یک مدل جدید برای سناریوسازی تقاضای دستگاه‌های خودپرداز (مورد مطالعه: دستگاه‌های خودپرداز شهر تهران). *آینده‌پژوهی مدیریت*، ۳۰(۳)، ۱۷۵–۱۸۸.

سنگبر، محمدعلی؛ صافی، محمدرضا؛ آذر، عادل و ربیعه، مسعود (۱۴۰۰). ارائه چارچوبی کمّی برای نگاشت شناختی فازی لایه‌ای با استفاده از رویکرد ترکیبی «نقشه خودسازمان‌دهنده» و «تئوری گراف و رویکرد ماتریس» (SOM-GTMA). *مدیریت صنعتی*، ۱۳(۱)، ۸۰– ۱۰۴.

عالم تبریز، اکبر؛ طلایی، حمیدرضا و مرادی، الناز (۱۳۹۲). ارزیابی عوامل کلیدی پیاده‌سازی موفق تولید در کلاس جهانی با استفاده از رویکرد یکپارچه مدل‌سازی ساختاری تفسیری (ISM)، تئوری گراف و رویکرد ماتریسی (GTMA) (مطالعه موردی: گروه ایران‌خودرو و سایپا). *مدیریت صنعتی*، ۵(۱)، ۶۳– ۸۱.

میرفخرالدینی، سیدحیدر و امیری، یاسر (۱۳۸۹). ارائه راه‌کارهای ارتقای خدمات الکترونیکی بانک‌ها با رویکرد BSC ،ANP فازی و TOPSIS فازی (مطالعه موردی: بانک‌های دولتی منتخب استان فارس). *مدیریت صنعتی*، ۲(۲)، ۱۴۱– ۲۹۸.

References

Abdulatif, A., Samarasinghe, R. & Thilakarathne, N. N. (2023). A Novel Robust Geolocation-Based Multi-Factor Authentication Method for Securing ATM Payment Transactions. *Applied Sciences*, 13(19), 10743.

Adamic, L. A. & Adar, E. (2003). Friends and neighbors on the web. *Social networks*, 25(3), 211-230.

Adesina, E., Adewuyi, A., Morenikeji, G., Ogundele, T. & Babatunde, F. (2022). Optimal Coverage Analysis of Existing Automated Teller Machines within Minna Metropolis, Nigeria using the Best-Fit Model. *International Journal of Environment and Geoinformatics*, 9(1), 127-139.

Agha Gholizadeh Sayyar, A., Motadel, M. & Pour ebrahimi, A. (2019). Presenting a new model for ATM demand scenario. *Journal of future studies management*, 30 (3), 175-188. (*in Persian*)

Alem Tabriz, A., Talaie, H. R. & Moradi, E. (2013). Evaluating the Key Factors of Successful Implementation of World Class Manufacturing Using an Integrated Approach of Interpretive Structural Modeling (ISM), Graph Theory and Matrix Approach (GTMA): A Case Study for Iran Khodro and Saipa in Iran. *Industrial Management Journal*, 5(1), 63-81. doi: 10.22059/imj.2013.35683 (*in Persian*)

Allamanis, M., Scellato, S. & Mascolo, C. (2012, November). Evolution of a location-based online social network: analysis and models. *In Proceedings of the 2012 internet measurement conference* (pp. 145-158).

Barabasi, A. L. (2005). The origin of bursts and heavy tails in human dynamics. *Nature*, 435(7039), 207-211.

Berahmand, K., Nasiri, E., Rostami, M. & Forouzandeh, S. (2021). A modified DeepWalk method for link prediction in attributed social network. *Computing*, 103, 2227-2249.

Brockmann, D., Hufnagel, L. & Geisel, T. (2006). The scaling laws of human travel. *Nature*, 439(7075), 462-465.

Clauset, A., Moore, C. & Newman, M. E. (2008). Hierarchical structure and the prediction of missing links in networks. *Nature*, 453(7191), 98-101.

Engerman, D. C. (2007). Bernath lecture: American knowledge and global power. *Diplomatic History*, 31(4), 599-622.

Gurgul, H. & Suder, M. (2018). Impact of ATM location on its profitability in Malopolskie and Podkarpackie provinces. *Managerial Economics*, 19(1).

Heckerman, D., Meek, C. & Koller, D. (2004). *Probabilistic models for relational data. Technical Report MSR-TR-2004-30*, Microsoft Research.

Huang, Z., Li, X. & Chen, H. (2005, June). Link prediction approach to collaborative filtering. *In Proceedings of the 5th ACM/IEEE-CS joint conference on Digital libraries* (pp. 141-142).

Jog, V. V. & Pardeshi, N. R. (2014). Advanced security model for detecting frauds in ATM transaction. *International Journal of Computer Applications*, 95(15), 47-50.

Lei, C. & Ruan, J. (2013). A novel link prediction algorithm for reconstructing protein–protein interaction networks by topological similarity. *Bioinformatics*, 29(3), 355-364.

Li, H., Ge, Y., Hong, R. & Zhu, H. (2016, August). Point-of-interest recommendations: Learning potential check-ins from friends. *In Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 975-984).

Li, Y. & Kettinger, W. J. (2022). Testing the relationship between information and knowledge in computer-aided decision-making. *Information Systems Frontiers*, 24(6), 1827-1843.

Liben-Nowell, D. & Kleinberg, J. (2003, November). The link prediction problem for social networks. *In Proceedings of the twelfth international conference on Information and knowledge management* (pp. 556-559).

Lü, L. & Zhou, T. (2011). Link prediction in complex networks: A survey. *Physica A: statistical mechanics and its applications*, 390(6), 1150-1170.

Mirfakhraddini, S. H. & Amiri, Y. (2010). Proposing solutions to improve E-banking services using BSC, FANP & FUZZY TOPSIS (Case study: Selected banks in Fars province). *Industrial Management Journal*, 2(2), 141-298. (*in Persian*)

Mwatsika, C. (2016). Factors influencing customer satisfaction with ATM banking. *International Journal of Academic Research in Business and Social Sciences*, 6(2), 26-41.

Narayanan, M. & Cherukuri, A. K. (2016). A study and analysis of recommendation systems for location-based social network (LBSN) with big data. IIMB Management Review, 28(1), 25-30.

Nasiri, E., Berahmand, K., Samei, Z. & Li, Y. (2022). Impact of centrality measures on the common neighbors in link prediction for multiplex networks. *Big Data*, 10(2), 138-150.

Nazari-Ganje, N. & Mirzapour Al-E Hashem, S. M. J. (2020). An integrated location-inventory routing problem for ATMs in banking industry: a green approach. *Modeling and Optimization in Green Logistics*, 27-52.

Newman, M. E. (2001). Clustering and preferential attachment in growing networks. *Physical review E*, 64(2), 025102.

Pao, H. K., Fadlil, J., Lin, H. Y. & Chen, K. T. (2012). Trajectory analysis for user verification and recognition. *Knowledge-Based Systems*, 34, 81-90.

Rastogi, A., Sharma, Y., Mukherji, S., Kaliyar, R. K. & Baghel, V. K. (2023, August). Predictive Analysis of Optimal Automated Teller Machine Site Selection Using Machine Learning and Deep Learning: A Comprehensive Study on Variables, Challenges, and Opportunities. *In 2023 International Conference on Electrical, Electronics, Communication and Computers (ELEXCOM)* (pp. 1-6). IEEE.

Rukpakavong, W., Subsomboon, K. & Nilpanich, S. (2022). Mutual authentication for cardless atm withdrawal using location factor. *Creative Science*, 14(2), 245396-245396.

Safdari, H., Contisciani, M. & De Bacco, C. (2022). Reciprocity, community detection, and link prediction in dynamic networks. Journal of Physics: Complexity, 3(1), 015010.

Sangbor, M. A., Safi, M. R., Azar, A. & Rabieh, M. (2021). Development a Quantitative Framework for Multilayer Fuzzy Cognitive Maps by combining "Self-Organizing Map" and "Graph Theory and Matrix Approach" (SOM-GTMA). *Industrial Management Journal*, 13(1), 80-104. doi: 10.22059/imj.2021.308177.1007769 (*in Persian*)

Suwirya, I. P., Candiasa, I. M. & Dantes, G. R. (2022). Evaluation of ATM Location Placement Using the K-Means Clustering in BNI Denpasar Regional Office. *Journal of Computer Networks, Architecture and High-Performance Computing*, 4(2), 158-168.

Takenova, K. & Guleva, V. Y. (2023). Determination of Optimal Locations for ATM Network Service Points. *Procedia Computer Science*, 229, 198-207.

Taskar, B., Abbeel, P., Wong, M. F. & Koller, D. (2007). Relational markov networks. *Introduction to statistical relational learning*, 175, 200.

Trang, P. T., Sonb, N. L. N. & Giangc, P. T. (2019). The Influence of ATM location characteristics on ATM usage in Vietnam. *International Journal of Advanced Engineering and Management Research* 4(03).

Wei, X., Liu, Y., Sun, J., Jiang, Y., Tang, Q. & Yuan, K. (2023). Dual subgraph-based graph neural network for friendship prediction in location-based social networks. *ACM Transactions on Knowledge Discovery from Data*, 17(3), 1-28.

Yuan, J., Zheng, Y., Xie, X. & Sun, G. (2011, August). Driving with knowledge from the physical world. *In Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 316-324).

Yuana, A., Terkimbi, T.E., Ortwer, A.B., Peter, A., Mwuese, G.G. & Inya, O.J. (2022). Isolation and Identification of Bacteria on Automated Teller Machines (ATMs) in Makurdi Metropolis. *Frontiers in Environmental Microbiology*, 8(1), 1-5. https://doi.org/10.11648/j.fem.20220801.11

Zhang, Z., Li, D., Song, Z., Duffield, N. & Zhang, Z. (2023, November). Location-Aware Social Network Recommendation via Temporal Graph Networks. *In Proceedings of the 7th ACM SIGSPATIAL Workshop on Location-based Recommendations*, Geosocial Networks and Geoadvertising (pp. 58-61).

Zheng, Y., Zhang, L., Ma, Z., Xie, X. & Ma, W. Y. (2011). Recommending friends and locations based on individual location history. *ACM Transactions on the Web (TWEB)*, 5(1), 1-44.