اولین کنفرانس مهندسی فناوری اطلاعات مکانی
دانشگاه صنعتی خواجه نصیرالدین طوسی
دانشکده مهندسی نقشه‌برداری
۲۹ و ۳۰ دی ماه ۱۳۹۴

The 1st National Conference on
Geospatial Information Technology
K.N.Toosi University of Technology
Faculty of Geomatics Engineering
19 - 20 January 2016

استفاده از مکان‌یابی سلسله مراتبی برای بهبود پیش بینی مکان آتی

رضا بکتاش[۱]، سامان هراتی زاده[۲*]،

۱- کارشناسی ارشد دانشکده علوم و فنون نوین دانشگاه تهران
۲- استادیار دانشکده علوم و فنون نوین دانشگاه تهران

چکیده :

در سال‌های اخیر، تحلیل و کاوش داده‌های مکانی جمع‌آوری شده از سنسورهای تلفن همراه، اهمیت زیادی در حوزه خدمات فراگیر شهری پیدا کرده‌است. یکی از چالش‌های مهم در این حوزه، پیش‌بینی جابه‌جایی و مکان آتی کاربران تلفن همراه است که می‌تواند در کاربردهای متفاوتی از جمله پیش بینی و مدیریت ترافیک و ارائه خدمات و تبلیغات هوشمندانه به کار رود. متاسفانه، با وجود تلاش‌های انجام شده در جهت ارتقاء سیستم‌های پیش‌بینی مکان، این سامانه‌ها همچنان با مشکل کمی دقت روبرو هستند. در این مقاله، سیستم نوینی برای پیش‌بینی مکان آتی کاربر ارائه می‌دهیم که با استفاده از روش‌های مکان‌یابی هرمی و توسعه الگوریتم‌های دنباله‌کاوی سازگار با ساختار هرمی ،دقت سامانه‌های فعلی را ارتقاء می‌دهد.

واژه‌های کلیدی : پیش بینی مکان آتی، خوشه بندی سلسله مراتبی، داده‌کاوی مکانی، دنباله کاوی.

نویسنده مکاتبه کننده : سامان هراتی زاده
آدرس پستی : تهران، امیرآباد، دانشکده علوم و فنون نوین، آزمایشگاه شماره ۲
تلفن : ۹۸۹۳۸۳۷۷۸۰۸۶
آدرس پست الکترونیک : rezabaktash@ut.ac.ir

۱- مقدمه

در سال‌های اخیر، با گسترش تلفن‌های همراه و تجهیز آن‌ها به سنسورهای مختلف، اطلاعات و داده‌های مختلفی درباره رفتار کاربران تلفن‌ها و دستگاه‌های همراه هوشمند فراهم کرده‌است. یکی از مهم‌ترین انواع داده‌های جمع آوری شده، داده موقعیت جغرافیایی است که با پردازش، تحلیل و استخراج الگو از آن‌ها، می‌توان اطلاعات بسیار مفیدی از کاربران دستگاه‌های همراه به دست آورد. از این اطلاعات می‌توان در سیستم‌های توصیه‌گر، شبکه‌های اجتماعی، کاربری‌های شهری و خدمات همراه، برای تبلیغات هوشمند، توصیه افراد و مکان‌ها به یکدیگر[1]، پیش‌بینی روند ترافیک شهری[2] و سایر خدمات فراگیر استفاده کرد. اکثریت این خدمات مبتنی بر موقعیت[1]، نیازمند سامانه‌ای هستند که یتواند موقعیت آتی کاربران را پیش‌بینی کند.

سیستم پیش‌بینی موقعیت آتی کاربران، با دریافت اطلاعات مکانی کاربران، تحلیلو استخراج الگوی حرکتی کاربران، موقعیت آن‌ها را در آینده پیش‌بینی می‌کنند. این سامانه‌ها، عموما، ابتدا داده‌های خام موقعیت را به نقاط توقف و سپس مکان‌های مهم تبدیل می‌کنند. برخی با استفاده از خوشه بندی نقاط موقعیت[5]–[3] و برخی با تقسیم بندی فضا[۶], [۷] مکان‌های مهم یا نقاط پرطرفدار[2] را تشکیل می‌دهند. سپس با استفاده از همین نقاط پرطرفدار، مدل خود را می‌سازند. روش‌های مختلفی برای ساخت مدل و یادگیری الگوی جابه‌جایی کاربر ارایه شده که قالبا مبتنی بر روش‌های احتمالاتی[9] ,[8], [6], [4], [3] و یا دنباله‌کاوی[۵], [۷], [۱۰], [۱۱] است.

مشکلات اصلی سامانه‌های فعلی پیش‌بینی موقعیت آتی را می‌توان از دو جهت تشخیص مکان و یادگیری الگوی حرکتی کاربران بررسی کرد. الگوریتم‌های فعلی تشخیص مکان است که نمی‌توانند مکان‌های کوچک و کم تکرار را شناسایی کنند. به همین دلیل، اطلاعات رفتاری کاربران در این مکان‌ها را از دست می‌دهند. روش‌های یادگیری الگو نیز با دو مشکل اصلی روبرو هستند: مدل‌های احتمالاتی، عموما از مدل مارکوف درجه ۲ استفاده می‌کنند و به خوبی نمی توانند الگوهای رفتاری موجود در دنباله طولانی حرکتی افراد را استخراج کنند. از طرف دیگر، روش‌های دنباله‌کاوی، در صورت قرار گرفتن کاربر در مکان کم تکرار، نمی‌توانند مکان آتی او را پیش‌بینی کنند. در این مقاله، روش جدیدی برای پیش‌بینی موقعیت آتی کاربران پیشنهاد می‌دهیم که مشکلات فوق را رفع می‌کند.

در این مقاله، ما سیستم جدیدی به نام LocArch برای پیش‌بینی مکان آتی کاربران ارائه می‌دهیم که با استفاده از یک الگوریتم تشخیص مکان هرمی، می‌تواند تمامی مکان‌ها در ریزدانگی‌های مختلف را شناسایی کند. سپس، با توسعه الگوریتم‌های دنباله‌کاوی، قادر است تا مکان کاربر در صورت قرار گرفتن در هر تمامی مکان‌ها، اعم از پرتکرار و نادر پیش‌بنی کند. به صورت خلاصه، مهم‌ترین نوآوری‌های این مقاله عبارتند از:

- ارائه یک سیستم نوین جهت پیش‌بینی مکان آتی کاربران
- ارائه یک روش تشخیص مکان با استفاده از خوشه‌بندی هرمی که بتواند تمامی مکان‌ها را شناسایی کند.
- ارائه یک روش پیش‌بینی مکان مبتنی بر دنباله‌کاوی که با ساختار هرمی سازگار باشد و بتواند در اکثریت مواقع، مکان کاربر را پیش‌بینی کند.

در ادامه این مقاله، بخش‌ها به این ترتیب خواهند بود. بخش ۲ به مرور کارهای مرتبط اختصاص دارد. در بخش ۳ به معرفی ساختار کلی سیستم پیش‌بینی و شرح جزئیات آن خواهیم پرداخت. در بخش ۴ نیز نتایج آزمایشات این سیستم پیش‌بینی بر روی داده‌های واقعی را ارایه خواهیم داد.

[1] Location-Based Services
[2] Points of Interest (POI)

۲- کارهای مرتبط

در سایر کارهای مرتبط با پیش‌بینی مکان آتی، از روش‌های مختلفی برای یافتن مکان‌های مهم استفاده شده است. آن‌ها نقاط توقف کاربران را به عنوان داده‌های ورودی در نظر گرفته و خوشه‌هایی که از تجمع این نقاط در برخی مکان‌های شهری تشکیل می‌شود را به عنوان مکان‌های مهم کاربر در نظر می‌گیرند. دسته‌ای از مقالات از انواع الگوریتم‌های خوشه‌بندی مانند خوشه‌بندی چگالی-محور[۳]–[۵] و خوشه‌بندی کا-میانگین[۳][۸] برای یافتن مکان‌های مهم استفاده کرده‌اند. برخی دیگر از الگوریتم‌های توری[۶], [۷] برای یافتن مکان‌های مهم بهره جسته‌اند. این الگوریتم‌ها با سلول‌بندی فضا و شمارش تعداد نقاط هر سلول، سلول‌های پرتراکم را تجمیع نموده و مکان‌های مهم را تشکیل می‌دهند. در مقاله [۱۰] علاوه بر خوشه‌بندی چگالی-محور در یک سطح، چند سطح دیگر از ترکیب خوشه‌ها تشکیل داده‌اند که برای شباهت کاربران از آن‌ها استفاده کرده‌اند.

یکی از مدل‌های استفاده شده در سیستم‌های پیش‌بینی، مدل احتمالاتی مارکوف است[۳], [۴], [۶], [۸], [۹]. در این مدل، مکان‌های شهری تبدیل به وضعیت‌های[۴] مختلف سیستم می‌شوند. سپس با استفاده از داده‌های آموزش موجود، احتمالات گذر از هر وضعیت به دیگری (با توجه به درجه مدل) محاسبه می‌شود. هر چه درجه مدل مارکوف بالاتر باشد، زمان آموزش آن بیشتر می‌شود. تعدادی از این مدل‌های احتمالی از مدل مارکوف مخفی استفاده کرده‌اند. این مدل با تعدادی وضعیت مخفی اضافی، وضعیت‌های مختلف مدل را به هم وابسته می‌کند. بعضی مقالات دیگر از سری زمانی غیرخطی به عنوان مدل برای پیش‌بینی استفاده می‌کنند[۱۲], [۱۳]. این مقالات، سری‌های زمانی را مستقل از هم برای هر مکان و با استفاده از زمان‌های ورود و مدت توقف، تشکیل می‌دهند. برای پیش‌بینی، زمان مورد نظر خود را به عنوان ورودی می‌دهند و مکان‌هایی را که در آن زمان، برای شخص مورد نظر، محتمل‌تر باشند را باز می‌گردانند. بسیاری از مقالات، مدل یادگیری خود را با استفاده از توالی‌کاوی (یافتن توالی‌های رایج) ساخته‌اند[۷], [۱۰], [۱۱], [۵]. در مقاله‌ی چن[۱۱]، داده‌ها را در ساختاری به نام درخت الگو، ذخیره می‌کنند و با استفاده آن، توالی‌های پرتکرار را استخراج می‌کنند. در مقاله ینگ[۵]، نیز درختی پیشوندی مشابه قبلی ساخته شده است.

۳- سیستم پیش‌بینی LocArch

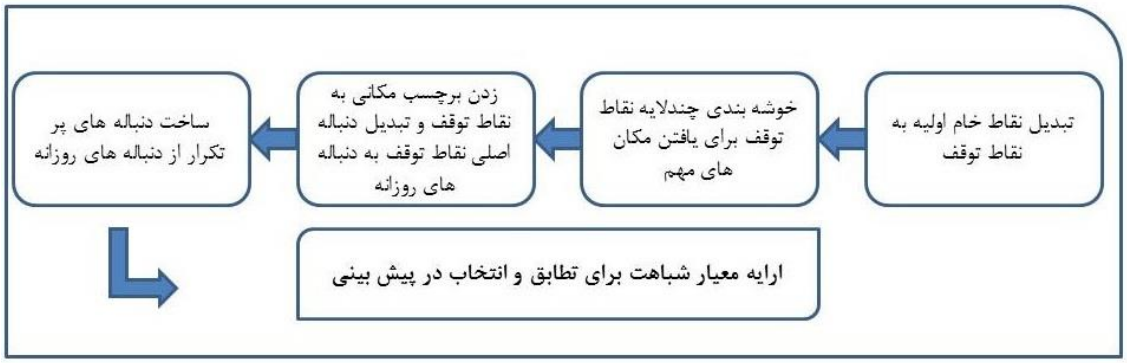
ما در سیستم LocArch مکان‌های مهم را در چند لایه ارایه کرده‌ایم که مبتنی بر خوشه‌بندی سلسله‌مراتبی بوده و در لایه‌های مختلف خود، مکان‌هایی با درشت‌دانه‌گی مختلف تشکیل می‌دهد. این ساختار مکان‌یابی را بعدا، در مدل یادگیری خود که مبتنی بر دنباله‌کاوی بوده و با ساختار هرمی مکان‌ها سازگار شده، استفاده کرده‌ایم. الگوریتم ما برای دنباله‌کاوی مبتنی بر اپریوری بوده و پس از ساخت کاندیدا به صورت ترتیبی دنباله‌های متداول به طول‌های مختلف را از شمارش کاندیدای ساخته شده و انتخاب کاندیدای متداول تشکیل می‌دهد.

هنگامی که مکان‌ها در یک سطح ساخته شده باشند، همواره با افزایش شعاع مکان‌ها، الگوهای ساخته شده کلی‌تر خواهند شد و خطای پیش‌بینی بیشتر. اما از طرفی با ریزدانه تر شدن مکان‌ها، الگوهای کمتر متداول ساخته می‌شوند که چه بسا سامانه پیش‌بینی، در بسیاری مواقع قادر به پیش‌بینی نباشد. با ساختن مکان‌ها در چند لایه، این دو ایراد به نوعی تقلیل داده می‌شوند. زیرا اولا در صورت وجود چند الگو در سطح‌های مختلف، اولویت را به تطابق با الگوهای سطوح پایین‌تر می‌دهیم و ثانیا در صورتی که در سطوح پایین تطبیق اتفاق نیفتد، در سطوح بالاتر اتفاق خواهد افتاد.

در این بخش، ساختار کلی سیستم پیش‌بینی مکان آتی کاربران را به طور خلاصه آورده‌ایم. سپس به بررسی هریک از بخش‌های سیستم پیش‌بینی LocArch پرداخته‌ایم

[3] K-Means
[4] State

سیستم ما در ابتدا داده‌های خام ثبت شده از دستگاه‌های گیرنده جی‌پی‌اس همراه را، که شامل موقعیت جغرافیایی و زمان و تاریخ دقیق بوده و برای هر کاربر به صورت جداگانه وجود دارد، را دریافت می‌کند و در پایان، با گرفتن دنباله‌ی حرکتی از یک کاربر، مکان بعدی حضور وی را پیش‌بینی می‌نماید.



شکل 1 - مراحل سیستم پیش‌بینی LocArch

مراحل سیستم پیش‌بینی را به ترتیب در اینجا آورده‌ایم. از بین آن‌ها، بخش‌هایی که نیاز به توضیح و بسط بیشتری دارند، به طور جداگانه آورده خواهد شد.

- **تبدیل نقاط خام اولیه به نقاط توقف:** نقاط خام اولیه شامل تعداد زیادی داده لحظه‌ای است که به طور متناوب ثبت شده است. منظور از نقاط توقف، محل‌هایی است که کاربر طی زمان زیادی (مثلا ۳۰ دقیقه) در شعاع محدودی (مثلا ۲۰۰ متر) سپری کرده است. با این تبدیل، حجم داده‌ها به طور چشمگیری کاهش می‌یابد و به حالت دنباله‌ای از نقاط در می‌آید [۱۰].

- **خوشه‌بندی چندلایه نقاط توقف برای یافتن مکان‌های مهم:** برای یافتن مکان‌های مهم کاربران، نقاط توقف آن‌ها را خوشه‌بندی کرده‌ایم. ما در این مرحله، برای اولین بار در سیستم‌های پیش‌بینی، از الگوریتم خوشه‌بندی سلسله مراتبی و از نوع پایین به بالا[5] استفاده کردیم. این الگوریتم در پایین‌ترین سطح، هر نقطه را یک خوشه در نظر می‌گیرد. سپس در هر مرحله حرکت به سمت بالا، نزدیک‌ترین خوشه‌ها را با هم ترکیب می‌کند. سپس با در نظر گرفتن برش‌هایی از سطوح ساخته شده، مکان‌های کاربر تعیین می‌شود. در ادامه توضیحات بیشتری درباره این الگوریتم خواهیم داد.

- **زدن برچسب مکانی به نقاط توقف و تبدیل دنباله اصلی نقاط توقف به دنباله‌های روزانه:** نقاط توقف اولیه در این مرحله با استفاده از برچسب‌هایی که در خوشه‌بندی به‌دست آوردند، برچسب زده می‌شوند. سپس دنباله‌ی اصلی تبدیل به دنباله‌های روزانه می‌شوند تا برای ساخت الگوهای متداول و پیش‌بینی آماده شوند. نحوه‌ی تبدیل دنباله‌ها، برچسب زدن مکانی و فیلتر کردن دنباله‌ها، در ادامه شرح داده خواهد شد.

- **ساخت دنباله‌های پرتکرار از دنباله‌های روزانه:** به منظور استخراج الگوی حرکتی افراد، لازم است تا دنباله‌های پرتکرار را از روی دنباله‌های گذشته کاربر استخراج نماییم. این کار با الگوریتم مبتنی بر اپریوری - یعنی ساخت کاندیدا و شمارش آن‌ها- انجام می‌شود. با این تفاوت که به خاطر وجود سلسله‌مراتب روی

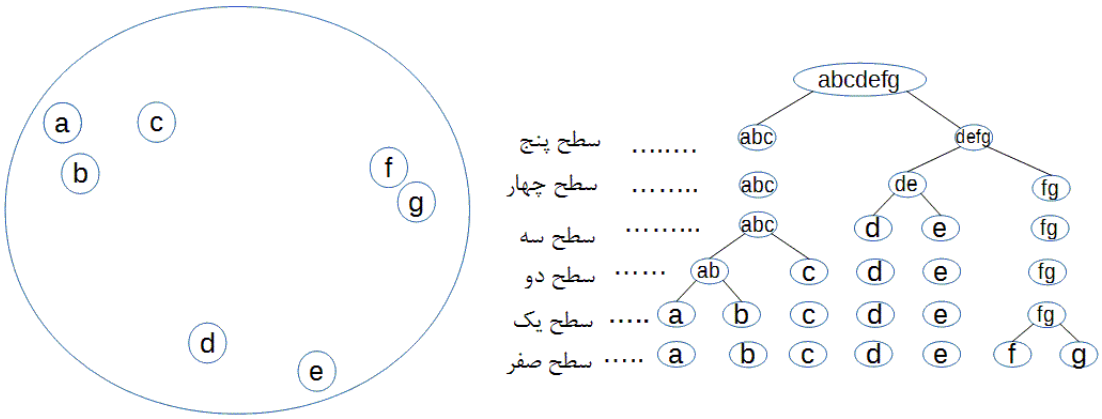
[5] Bottom-Up

مکان‌های موجود، هر مکان دارای روابطی با مکان‌های کوچکتر و بزرگتر از خود است. این موضوع تغییراتی در در الگوریتم به وجود می‌آورد که در ادامه به آن‌ها خواهیم پرداخت.

- **ارایه معیار شباهت برای تطابق و انتخاب در پیش‌بینی:** سیستم پیش‌بینی مکان آتی، برای پیش‌بینی موقعیت کاربر در دنباله جدید و فعلی، شباهت آن را با دنباله‌ها پرتکرار موجود در سیستم محاسبه می‌کند. سپس، از میان دنباله‌های متداول، آن‌هایی که شباهت بیشتری به دنباله قعلی دارند، انتخاب شده و از روی آن‌ها رای گیری برای آیتم بعدی انجام می‌شود. در این بین، انتخاب دنباله‌های متداول از سطوح مختلف هم دارای اولویت‌بندی خواهد بود.

۳-۱- مکان‌یابی با خوشه‌بندی سلسله‌مراتبی

LocArch برای یافتن مکان‌های شهری، از یک روش هرمی پایین به بالا بهره می‌برد. در این روش، ابتدا هریک از نقاط به عنوان یک خوشه در نظر گرفته می‌شوند. سپس طی هر مرحله دو تا نزدیک‌ترین خوشه‌ها با هم پیوند داده شده و یکی می‌شوند. این کار تا زمانی ادامه می‌یابد که فقط یک خوشه باقی بماند. حال با توجه به مساله و نیاز آن، سطح یا سطوحی از خوشه‌ها انتخاب می‌شود (شکل ۱). هر خوشه C_i دارای لیستی از نقاط $members$ و اندیس $index$ است.



شکل 2خوشه‌بندی سلسله مراتبی پایین به بالا

در ابتدای الگوریتم به تعداد نقاط توقف خوشه داریم که هرکدام یک عضو دارند. فاصله دو خوشه را از فرمول زیر که میانگین فاصله‌های اعضای دو خوشه است محاسبه می‌کنیم که در آن، $|p-p'|$به معنی فاصله اقلیدسی دو نقطه p' و p است.

$$Average_Dist(C1,C2)=\frac{1}{n1.n2}\sum_{p\in C1}\sum_{p'\in C2}|p-p'|$$

همان‌طور که گفته شد، در هر مرحله از خوشه‌بندی فاصله همه خوشه‌ها با یکدیگر محاسبه شده و از بین آن‌ها، آن زوج خوشه‌هایی که کمترین فاصله را دارند با هم ترکیب می‌شوند. این فاصله کمینه در هر سطح i به نام $(minz_i)$ در مکان‌یابی استفاده می‌شود. فرض کنید تعداد سطح‌های در نظر گرفته شده برای برش، k باشد. قبل از خوشه‌بندی یک لیست از مقادیر آستانه[6] به طول k در نظر می‌گیریم. در طی خوشه‌بندی هرگاه فاصله $minz_i$ بیش از آستانه کنونی شود، خوشه‌های سطح i ام به عنوان یکی از برش‌های دلخواه انتخاب می‌شوند.

[6] Cut-off list

```
Procedure Hierarchical-Clustering
CL = list of clusters where each cluster has one stay point
N = number of stay points
Map<Int,Cluster> result;
cut-offs = list of thresholds
for(i=0 ; i<N ; i++)
<CL,minz[i]> = level-up(CL);
if(minz[i]>cut-offs.first())
        result.add(<i,CL>);
end if
end

function level-up(Cluster_List CL){
        double[][] distances = distances between each pair of clusters
min i,j  = cluster i and j which have shortest distance;
remove i and j from CL;
C_new = merge(i,j);
add C-new to CL;
return CL;
}
```

شکل 3 - شبه کد الگوریتم خوشه‌بندی

در انتهای این الگوریتم، خوشه‌هایی که هر نقطه توقف در سطوح مختلف متعلق به آن‌ها است، مشخص می‌شوند. به این ترتیب برچسب‌های مکانی نقاط توقف، که لیستی مرتب هستند، زده می‌شوند.

همان‌طور که در شکل ۲ نشان داده‌شده است، خوشه‌های به‌دست آمده به خاطر روش ساخت خود از یک ساختار درختی پیروی می‌کنند. ما در این مرحله، فقط با استفاده از سطوح منتخب خوشه‌ها درختی می‌سازیم. این درخت بعدا در استخراج توالی‌های پرتکرار استفاده می‌شود. در درخت ساخته شده، هریک از راس‌ها (خوشه‌ها) دارای یک پدر بوده و یک راس مشترک به نام ریشه پدر راس‌های بالاترین سطح خواهد بود.

۳-۲- زدن برچسب مکانی به نقاط توقف و تبدیل دنباله اصلی نقاط توقف به دنباله‌های روزانه: در این مرحله، داده‌های موجود که دنباله‌ی نقاط توقف هستند، برچسب مکانی زده می‌شوند. همان‌طور که در مرحله خوشه‌بندی دیدیم، هر نقطه توقف می‌تواند در سطح‌های مختلف و حداکثر به تعداد کل سطح‌ها، عضو خوشه‌های مختلف باشد. بنابراین هر نقطه توقف به تعداد سطح‌های خود برچسب خواهد داشت و از آنجا که ترتیب این

شناسه‌های مکانی هم اهمیت دارد، این شناسه‌ها در یک لیست مرتب به نام شناسه سطوح ذخیره می‌شود. سپس، لیست اصلی نقاط توقف را با استفاده از اطلاعات زمانی نقاط توقف به دنباله‌های روزانه تبدیل می‌کنیم. در این مرحله، لازم است تا دنباله‌های روزانه فیلتر شود زیرا در مواردی نقاط توقف متوالی دارای برچسب یکسان هستند و نیاز است در این موارد یکی از دو نقطه متوالی حذف شود. هم‌چنین دنباله‌های روزانه‌ای که در انتهای فیلترینگ، طولشان کمتر از ۲ شود، از میان دنباله‌های روزانه حذف می‌شوند.

۳-۳- استخراج دنباله‌های متداول

۳-۳-۱- تعاریف

الف) دنباله‌ی نقاط توقف: مجموعه ای مرتب از نقاط توقف S که دارای موقعیت مکانی و زمان شروع و پایان و نیز لیستی از برچسب‌های مکانی است. $S = \{s_1, s_2, \dots, s_n\}$

ب) توالی: مجموعه‌ای مرتب از آیتم‌ست‌ها است که هرکدام شامل یک یا چند آیتم هستند. ترتیب زمانی آیتم‌ست‌ها در توالی مهم است، اما آیتم‌های موجود در یک آیتم‌ست در یک زمان رویت شده‌اند.

$Sequence = \{ItemSet_1, ItemSet_2, \dots, ItemSet_n\}$

$ItemSet = \{item_1, item_2, \dots, item_m\}$

ج) وجود یک توالی در دیگری با مقدار حداکثر فاصله: توالی $S_m = \{is_1, is_2, \dots, is_m\}$ در توالی $S_n = \{is'_1, is'_2, \dots, is'_n\}$ وجود دارد، هرگاه m کوچکتر از n باشد و اندیس‌هایی $1'$ تا m' موجود باشند به طوری که $is_1 \subseteq is'_1$، و ... و $is_m \subseteq is'_m$ و این در حالی است که فاصله‌ی هر دو اندیس i' و j' از مقدار حداکثر فاصله[7] نباید بیشتر باشد.

۳-۳-۲- الگوریتم استخراج

دنباله‌های پرتکرار طبق الگوریتم کاوش دنباله های پرتکرار مبتنی بر اپریوری که قبلا اشاره شد ساخته می‌شوند. پارامتر «حداقل مقدار پشتیبان»[8] به عنوان ورودی این الگوریتم داده می‌شود و ساخت کاندیدا و شمارش آنها برای طول های مختلف دنباله انجام می‌شود.

این الگوریتم ابتدا عناصر یکه را شمارش می‌کند و در واقع دنباله‌های کاندید به طول یک را می‌سازد؛ سپس کاندیدا را شمارش می کند و آنهایی که تکرارشان در روزهای مختلف از عدد حداقل مقدار پشتیبان بیشتر باشند، به عنوان دنباله‌های متداول به طول یک انتخاب می‌شوند. سپس از ترکیب دنباله‌های متداول به طول یک، کاندیدای به طول دو را می‌سازد و دفعات وجود آن‌ها را در دنباله‌های روزانه کاربر –با توجه به تعریف ج- شمارش می‌کند. این‌بار دنباله‌های متداول به طول دو ساخته می‌شود. این کار تا زمانی ادامه می‌یابد که در یک مرحله نتوان هیچ کاندیدی ایجاد کرد

[7] max_gap
[8] min_sup

```
procedure Generate Frequent Sequences

input: list of daily stay point sequences

frequent_sequences = empty
max_len = maximum length of sequences

C1 = generate all single item sequences

L1 = items in C1 where count of their appearance in data is larger than min_sup

add L1 to frequent_sequences

for(i=2:max_len)

                    generate Ci from Li-1

                    generate Li from Ci

if Li is empty break

                    add Li to frequent_sequences

end
```

شکل 4-شبه کد الگوریتم استخراج دنباله‌های متداول

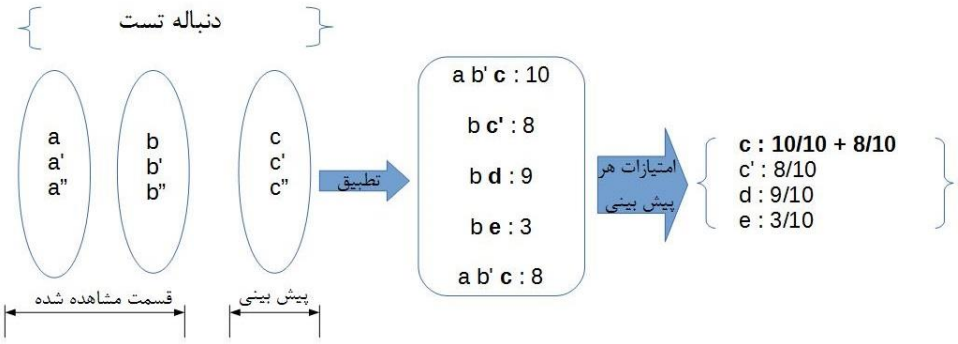
۳-۳-۳- هرس کردن دنباله‌های متداول:

با توجه به چند لایه بودن مکان‌ها، ممکن است حالت‌هایی در دنباله‌ها به وجود بیاید که از نظر معنایی دارای اشکال باشد. با توجه به درخت خوشه‌ها، مثلا اگر در یک دنباله متداول دو عنصر متوالی باشندکه اولی در مسیر راس تا ریشه درخت با دومی برخورد داشته باشد به این معناست که عنصر دوم از اجداد اولی است. حالت برعکس این هم ممکن است اتفاق بیافتد. ما در زمان ساخت دنباله‌های متداول از قرار گرفتن چنین عناصری پشت سر هم جلوگیری می‌کنیم. مثلا فرض کنید در یک دنباله‌ای داشته باشیم: {خانه، فروشگاه، دانشگاه، رستوران، خانه} و فرض کنید خانه و فروشگاه به دلیل نزدیک بودن، در سطح دوم خوشه‌بندی دارای اندیس‌های مکانی یکسان در محدوده محله باشند. یکی از کاندیدایی که برای شمارش ساخته می‌شود {خانه، محله} است. اما از طرفی خانه هم جزیی از محله است و این دنباله از نظر معنایی در سیستم ما مورد قبول نیست و حذف خواهد شد.

۳-۴- پیش‌بینی مکان بعدی

در صورت وجود دنباله جدید، سیستم لازم است تا شباهت آن را با دنباله‌های متداول محاسبه کند. معیار شباهت در حالت کلی پسوند مشترک میان دو دنباله است. از میان دنباله‌های متداولی که شباهت با دنباله کنونی دارند، آن‌هایی که در سطوح پایین‌تری تطبیق داده شده‌اند، با اولویت بیشتری انتخاب می‌شوند. هم‌چنین مقدار پشتیبان[9] دنباله‌ی متداول هم تاثیر مستقیم در امتیاز آن دنباله دارد. با این وجود، هریک از کاندیدا یک گزینه را به عنوان عنصر بعدی پیشنهاد می‌دهد. در نهایت مجموع امتیازات برای هر عنصر جمع می‌شود و عنصری که بیشترین امتیاز را داشته باشد، پیشنهاد داده می‌شود (شکل 5- **تطبیق الگو و انتخاب پیش‌بینی**).

[9] Support



شکل 5- تطبیق الگو و انتخاب پیش‌بینی

۴- ارزیابی و نتایج

در این قسمت، سیستم پیش‌بینی را با دادگان واقعی آزمایش کرده‌ایم. دادگان موجود به نام جئولایف[10] توسط مرکز تحقیقات مایکروسافت طی چهار سال و از ۱۸۲ کاربر در شهر پکن گردآوری شده است. ما از بین کاربران موجود، ۶۰ تا(حدود یک سوم) از کاربران که دارای بیشترین حجم داده بودند را بررسی و آزمایش کرده‌ایم. دنباله‌های روزانه هر کاربر در بخش‌های قبل به دو قسمت آموزش و آزمون تقسیم می‌شوند. از داده‌های آموزش در ساخت دنباله‌های متداول استفاده کردیم. داده‌های آزمون برای هر کاربر شامل تعدادی دنباله روزانه از نقاط توقف است. ما برای آزمایش سیستم خود، برای یک دنباله مانند $S = s_1, s_2, s_3, \dots, s_k$ از عنصر دوم به بعد همه عناصر را با در نظر گرفتن ماقبل خود به عنوان ورودی پیش‌بینی می‌کنیم. قسمت ماقبل دنباله به عنوان پیشینه با همه دنباله‌های متداول مقایسه می‌شود.

ارزیابی نتایج را از دو روش انجام داده‌ایم. اولین روش دقت پیش‌بینی مکان درست، با توجه به برچسب مکان نقاط توقف است. درصد دقت به‌دست آمده حاصل تقسیم تعداد پیش‌بینی‌های درست بر کل تعداد پیش‌بینی هاست. دومین روش محاسبه میانگین فاصله خطای مکان پیش‌بینی شده با محل واقعی نقطه توقف است.

$$accuracy = \frac{\#\ correct\ labels}{\#\ all\ predictions} \quad , \quad avg_error = \frac{\sum |l_{pr} - l_{real}|}{\#\ predictions}$$

برای مقایسه کارایی سیستم‌مان از یک سیستم دیگر با روش مکان‌یابی چگالی-محور (DBSCAN)[11] و توالی‌کاوی مبتنی بر اپریوری[12] استفاده کرده‌ایم. این سیستم برای خوشه‌بندی از الگوریتم خوشه‌بندی چگالی-محور با پارامترهای ۳۰ و ۳ به ترتیب برای «شعاع همسایگی» و «حداقل نقاط همسایه» استفاده کرده است. هم‌چنین از روش مبتنی بر اپریوری و استخراج قوانین انجمنی برای ساخت مدل پیش‌بینی استفاده کرده است. در روش خوشه‌بندی چگالی-محور نقاط نویز حذف می‌شوند و نمی‌توان در پیش‌بینی اندیس بر آن‌ها گذاشت. بنابراین برای معادل کردن خوشه‌بندی سلسله‌مراتبی با آن، نویزهای مربوط به چگالی-محور را حذف نموده ایم.

هم‌چنین در مواردی که برای تست هیچ دنباله متداولی مطابقت داده نشده، انتخاب تصادفی از بین مکان‌ها انجام گرفته است. مقایسه این دو آزمایش در جدول ۱ آمده است.

[10] GeoLife
[11] Density-based spatial clustering of applications with noise
[12] Apriori-Based

جدول ۱ مقایسه نتایج مکان‌یابی چگالی-محور و چندسطحی

سیستم پیش‌بینی	مکان‌یابی	تعداد نمونه	مقدار پشتیبان	درصد تطبیق	دقت پیش‌بینی (درصد)	میانگین خطای فاصله‌ای (متر)
اپریوری	DBSCAN	۱۲۱۶	۰.۱	۹۳.۸	۴۹	۲۷۳۷
LocArch	چند لایه	۱۲۳۱	۰.۱	۹۲.۶	۶۹	۲۰۲۰

آزمایش دیگری که انجام داده‌ایم (جدول 2) ، مقایسه مکان‌یابی سلسله‌مراتبی یک سطحی و چند سطحی در سیستم پیش‌بینی ما است. در این آزمایش، همه نقاط اولیه بدون حذف در مکان‌یابی حضور می‌یابند. به دلیل افزایش تعداد نمونه‌ها نسبت به جدول ۱ و با توجه به اینکه نقاط اضافه شده قبلا نویز بوده‌اند، خطا به مقدار زیادی افزایش یافته است.

جدول 2 - مقایسه نتایج مکان‌یابی یک سطحی و چند سطحی

روش مکان‌یابی	تعداد نمونه	مقدار پشتیبان	درصد تطبیق	دقت پیش‌بینی (درصد)	میانگین خطای فاصله‌ای (متر)
یک سطحی	۲۵۰۵	۰.۱	۴۵.۷	۲۰	۴۶۱۹
چند سطحی	۲۵۰۵	۰.۱	۶۹.۴	۳۷.۴	۴۸۸۴

۵- نتیجه گیری

در این مقاله، ما روش نوینی برای پیش‌بینی مکان آتی کاربران با استفاده از تحلیل داده‌های جمع‌آوری شده توسط سنسورهای تلفن همراه ارائه دادیم. سیستم پیشنهادی، با ارائه روشی برای تشخیص هرمی مکان‌ها و توسعه روش‌های دنباله‌کاوی، می‌تواند مکان آتی کاربر را در شرایط مختلف پیش‌بینی کند. لازم به ذکر است که چنین رویکردی، در الگوریتم‌های پیشین وجود ندارد و آن‌ها تنها می‌توانند مکانی آتی کاربر را در صورت قرار گرفتن در مکان‌های پر تراکم پی‌بینی کنند. نتایج آزمایش‌های کنونی، نشان می‌دهد که الگوریتم پیشنهادی به خوبی توانسته است به این هدف خود دست یابد و درصد تطابق بالاتری نسبت به الگوریتم‌های پیشین دارد. علاوه بر این، سیستم در شرایط مشابه وبا شرایط تطبیق یکسان، دقت بالاتری را نیز داشته‌اند.

مراجع

[1] Y. U. Zheng, L. Zhang, and X. Xie, “Recommending Friends and Locations Based,” vol. 5, no. 1, 2011.

[2] L. X. Pang, S. Chawla, W. Liu, and Y. Zheng, “On detection of emerging anomalous traffic patterns using GPS data,” *Data Knowl. Eng.*, vol. 87, pp. 357–373, 2013.

[3] J. Alvarez-lozano, “Learning and User Adaptation in Location Forecasting,” pp. 461–470, 2013.

[4] S. Gambs, M.-O. Killijian, and M. N. Del Prado Cortez, “Next place prediction using mobility Markov chains,” in *Proceedings of the First Workshop on Measurement Privacy and Mobility MPM 2012*, 2012, pp. 1–6.

[5] J. J. Ying, W. Lee, T. Weng, and V. S. Tseng, “Semantic trajectory mining for location prediction,” *Proc. 19th ACM SIGSPATIAL Int. Conf. Adv. Geogr. Inf. Syst. - GIS ’11*, p. 34, 2011.

[6] W. Mathew, R. Raposo, and B. Martins, “Predicting future locations with hidden markov models,” in *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, 2012, pp. 911–918.

[7] P. R. Lei, T. J. Shen, W. C. Peng, and I. J. Su, “Exploring spatial-temporal trajectory model for location prediction,” *Proc. - IEEE Int. Conf. Mob. Data Manag.*, vol. 1, pp. 58–67, 2011.

[8] D. Ashbrook and T. Starner, “Using GPS to learn significant locations and predict movement across multiple users,” *Pers. Ubiquitous Comput.*, vol. 7, no. 5, pp. 275–286, 2003.

[9] G. Gidófalvi and F. Dong, “When And Where Next: Individual Mobility Prediction,” *Proc. First ACM SIGSPATIAL Int. Work. Mob. Geogr. Inf. Syst. - MobiGIS ’12*, p. 57, 2012.

[10] Q. Li, Y. Zheng, X. Xie, Y. Chen, W. Liu, and W. Ma, “Mining User Similarity Based on Location History,” no. c, 2008.

[11] L. Chen, M. Lv, Q. Ye, G. Chen, and J. Woodward, “A personal route prediction system based on trajectory data mining,” *Inf. Sci. (Ny).*, vol. 181, no. 7, pp. 1264–1284, 2011.

[12] M. De Domenico, A. Lima, and M. Musolesi, “Interdependence and Predictability of Human Mobility and Social Interactions.”

[13] S. Scellato, M. Musolesi, C. Mascolo, V. Latora, and A. T. Campbell, “NextPlace : A Spatio-Temporal Prediction Framework for Pervasive Systems.”

Using Hierarchical Location Detection to Improve Next Location Prediction

Baktash, R.*[1], Haratizadeh, S. [2]

1-Ms.c student of Decision Science and Knowledge Engineering, Faculty of New Sciences and Technology, University of Tehran

2- Assistant professor in Faculty of New Sciences and Technology, University of Tehran

Abstract

Analyzing and mining of spatial data, gathered by mobile phone sensors, has gained attention in urban pervasive services recent years. One Important challenge in this area is future location prediction of mobile phone users, which has many applications specially in traffic prediction and intelligent advertisement and services. Unfortunately, in spite of many efforts for improvement of those systems, they still have low accuracy. In this paper, we present a new system LocArch for future location prediction, which uses hierarchical location detection and sequence mining algorithm, which is extended for hierarchical locations. Results of our experiments on real dataset show improvement in accuracy and lower error rate.

Keywords: future location prediction, hierarchical clustering, spatial data mining, sequence mining

Correspondence Address: KDD group, Faculty of New Sciences and Technology, University of Tehran, Tehran, Iran. **Tel: +98 938 377 8086**

Email: rezabaktash@ut.ac.ir