

تعیین کاربری اجتماعی در محیط‌های شهری بر مبنای بعد معنایی داده‌های شبکه‌های اجتماعی مکان‌مبنا

مرتضی طیبی^۱، فرید کریمی پور^{۲*}، کیمیا آموزنده^۱

^۱ دانشجوی کارشناسی ارشد سیستم‌های اطلاعات مکانی - دانشکده مهندسی نقشه‌برداری و اطلاعات مکانی -

پردیس دانشکده‌های فنی - دانشگاه تهران

{mortezaatayebi, amouzandeh.kimia}@ut.ac.ir

^۲ استادیار دانشکده مهندسی نقشه‌برداری و اطلاعات مکانی - پردیس دانشکده‌های فنی - دانشگاه تهران

fkarimipr@ut.ac.ir

(تاریخ دریافت مرداد ۱۳۹۶، تاریخ تصویب آبان ۱۳۹۶)

چکیده

شناخت محیط‌های شهری و درک رفتار حرکتی شهروندان یک زمینه تحقیقاتی مهم در حوزه تحلیل داده‌های مکانی است. شبکه‌های اجتماعی مکان‌مبنا داده‌های به‌روز، غنی و عظیمی را که کاربران به‌طور صادقانه‌ای از رفتار مکانی، زمانی و معنایی خود به اشتراک می‌گذارند، ثبت و جمع‌آوری می‌کنند. در این تحقیق تلاش شده است بر مبنای داده‌های مکانی و معنایی شبکه‌های اجتماعی مکان‌مبنا، محیط‌های شهری بر مبنای کاربری‌های اجتماعی خوشه‌بندی شوند. بدون تردید ساختار فیزیکی و همچنین کاربری یک منطقه شهری بر روی رفتار مکانی شهروندان تأثیر می‌گذارد و این تأثیر به داده‌های شبکه‌های اجتماعی مکان‌مبنا نیز سرایت می‌کند. در این راستا در اولین گام با به‌کارگیری یک الگوریتم خوشه‌بندی مبتنی بر شبکه عصبی رقابتی (SOM) داده‌های مکانی کاربران خوشه‌بندی می‌گردد. سپس با رسم دیاگرام ورونوی بر روی مراکز خوشه‌ها، محیط شهری به چند منطقه افراز شده و داده‌هایی که کاربران در هر منطقه ثبت کرده‌اند مشخص می‌شود. پس از استخراج درصد هر گروه از داده‌های معنایی و با در نظر گرفتن وزن هر یک از این گروه‌ها، با بهره‌گیری از الگوریتم فرا ابتکاری ژنتیک شاخص بهینه‌ای برای تعیین کاربری اجتماعی محاسبه می‌گردد. سپس با استفاده از یک الگوریتم خوشه‌بندی بر مبنای شاخص تعیین شده، به هر منطقه یک بعد معنایی که نماینده کاربری اجتماعی آن منطقه است، نسبت داده شد. به‌منظور ارزیابی روش پیشنهادی، نمودار تغییرات زمانی تعداد داده‌ها در طول شبانه‌روز برای پنج روز کاری هفته و دو روز پایان هفته برای هر گروه از داده‌های معنایی رسم شده و برای خوشه‌های شناسایی شده از روش پیشنهادی نیز استخراج گردید. سپس از همبستگی بین این نمودارها به‌عنوان شاخص ارزیابی روش پیشنهادی استفاده شد. نتایج حاصل از این تحقیق، بیانگر پتانسیل بالای شبکه‌های اجتماعی مکان‌مبنا برای شناخت محیط‌های شهری می‌باشد.

واژگان کلیدی: شبکه‌های اجتماعی مکان‌مبنا، خوشه‌بندی SOM، کاربری اجتماعی محیط‌های شهری، شاخص DB

* نویسنده رابط

۱- مقدمه

امروزه گسترش محیط‌های شهری و افزایش روزافزون جمعیت در کلان‌شهرها، از یک‌سو چالش‌های زیادی را با خود به همراه داشته است و از سوی دیگر سبب ایجاد منابعی نظیر گوشی‌های تلفن همراه، شبکه‌های اجتماعی، GPS خودروها و غیره برای تولید داده‌های مکانی انبوه و ارزشمند شده است. دولتمردان در راستای حل این چالش‌ها و اتخاذ تصمیم‌های حوزه مدیریت شهری نیاز به شناخت محیط‌های شهری و درک رفتار حرکتی شهروندان دارند که می‌تواند از منابع مذکور حاصل شود. محاسبات شهری^۱ اصطلاحی است که به این حوزه از مطالعات مکانی شهرها اختصاص داده شده است [۱]. این تحقیق به دنبال شناخت محیط‌های شهری و به‌طور ویژه تعیین کاربری در محیط‌های شهری می‌باشد. شناخت محیط‌های شهری می‌تواند از دو دیدگاه مورد بررسی قرار گیرد: نخست از دیدگاه ساختار فیزیکی که با منابع مختلفی از جمله سنجش‌ازدور حاصل می‌شود و دوم از دیدگاه کاربری اجتماعی که می‌تواند با بهره‌گیری از منابع متعددی مانند شبکه‌های اجتماعی مکان‌مبنا به‌دست آید. ساختار فیزیکی یا همان تعیین نوع پوشش زمینی (شهری، جاده، زمین کشاورزی، مراتع و جنگل‌ها و غیره) با روش‌هایی نظیر طبقه‌بندی، ادغام و غیره در سنجش‌ازدور حاصل می‌شود. حال اگر این شناخت با درک اجتماعی از محیط ترکیب شود می‌تواند نتایج با کارایی بالاتری را حاصل کند [۲]. به‌صورت کلی یک شبکه‌ی اجتماعی، بازسازی‌کننده‌ی روابط اجتماعی افراد در محیط‌هایی مانند یک وبگاه بوده و راه‌هایی را برای کاربران به منظور به اشتراک‌گذاری ایده‌ها، فعالیت‌ها، اتفاقات و علائق بر روی بستر اینترنت فراهم می‌کند. ظهور و فراگیر شدن گوشی‌های هوشمند و مجهز به GPS که به شبکه جهانی وب دسترسی دارند، موجب پیدایش شبکه‌های اجتماعی مکان‌مبنا گردیده است. شبکه‌های اجتماعی مکان‌مبنا افراد را قادر ساخته بعد مکانی را نیز به روش‌های مختلف، به شبکه‌های اجتماعی موجود اضافه کنند. برای مثال کاربرها می‌توانند عکس‌های خود را با مکان نشانه‌گذاری کرده و در شبکه‌های اجتماعی منتشر کنند (فلیکر^۲)، یا نظر خود را در مورد یک رویداد و

زمانی که در محل آن رخداد هستند بیان کنند (توییتر^۳). یا هر لحظه دوستانشان را از مکان خود مطلع کنند (فوراسکوئر^۴) [۳]. باوجود آنکه مدت زیادی از ظهور شبکه‌های اجتماعی مکان‌مبنا نمی‌گذرد، اما این شبکه‌ها هم در زمینه تجاری و هم در زمینه تحقیقاتی مورد توجه زیادی قرار گرفته و رشد چشمگیری داشته‌اند. در همین راستا، گروه تحقیقاتی شرکت مایکروسافت به سرپرستی ژنگ تحقیقات زیادی انجام داده‌اند. آن‌ها در سال ۲۰۰۸ با بررسی پیشینه مکانی کاربران میزان مشابهت افراد را مورد بحث قرار دادند [۴]. یکی دیگر از مهم‌ترین تحقیقات انجام شده توسط این گروه پروژه‌ی ژئولایف^۵ در سال ۲۰۰۹ بود که هدف آن ارائه یک شبکه اجتماعی مکان‌مبنا بود که بر مبنای داده‌های خط السیر GPS که در شبکه به اشتراک گذاشته می‌شد عمل نماید. از مهم‌ترین بخش‌های این سیستم می‌توان به مدل کردن داده‌های خط السیر کاربران در یک چارچوب سلسله مراتبی، کاوش جذاب‌ترین مکان‌ها در یک محدوده جغرافیایی و برنامه‌ریزی سفر برای کاربران و ارائه سرویس توصیه‌گر فردگرا اشاره کرد [۵]. در تحقیق دیگری در سال ۲۰۰۹ لی و همکاران برای مطالعه چگونگی به اشتراک‌گذاری داده‌های مکانی توسط کاربران، داده‌های زیادی از شبکه‌ی اجتماعی مکان‌مبنا برای تایت کیت^۶ جمع‌آوری کرده و نتایج تحلیل این داده‌ها از نظر ویژگی‌های کاربران، مشخصات حرکتی کاربران و ارتباط آن‌ها ارائه شد [۶]. همچنین اسکلاتو و همکاران در سال ۲۰۱۰ با تحلیل داده چهار شبکه اجتماعی مکان‌مبنا مختلف و با استفاده از یک معیار ابتکاری، فرضیه وابستگی ساختار اجتماعی کاربران به وضعیت جغرافیایی آن‌ها را بررسی کردند [۷]. در نهایت، کرانشو و همکاران در سال ۲۰۱۲ با استفاده از داده‌های شبکه‌های اجتماعی مکان‌مبنا و با استفاده از الگوریتم خوشه‌بندی به شناسایی مناطق شهری پرداختند [۸].

این تحقیق با بهره‌گیری از داده‌های مکانی، معنایی و زمانی کاربران در شبکه‌های اجتماعی، به خوشه‌بندی محیط‌های شهری در غالب مناطقی با کاربری‌های مختلف می‌پردازد. از تحقیقاتی که پیش از این در این زمینه

^۳ Twitter
^۴ Foursquare
^۵ GeoLife
^۶ Brightkite

^۱ Urban computing
^۲ Flickr

تحقیق است، اشاره کرد. به این نوع تحقیق و کاوش در محیط‌های شهری، اصطلاح "سنجش اجتماعی"^۱ اطلاق شده است. این اصطلاح از دو جهت با این نوع تحقیقات تناسب دارد: اولاً از نظر هم‌راستا بودن با سنجش از دور و ثانیاً از این جهت که هر شهروند در یک اجتماع نقش یک حس‌گر را ایفا می‌کند [۱۱]. جدول (۱) شباهت‌ها و تفاوت‌های سنجش از دور و سنجش اجتماعی را به شکل واضح‌تری بیان می‌کند. این تحقیق در تلاش است تا "کاربری اجتماعی"^۲ را در محیط‌های شهری تعیین کند. کاربری اجتماعی به نوع استفاده شهروندان از یک منطقه اطلاق می‌شود. به‌عنوان مثال منطقه‌ای با کاربری اجتماعی تفریحی نشان می‌دهد که شهروندان غالباً به منظور پرداختن به تفریحات خود به این منطقه سفر می‌کنند.

جدول ۱- مقایسه سنجش از دور و سنجش اجتماعی

جنبه مقایسه	سنجش از دور	سنجش اجتماعی
منبع داده	تصاویر ماهواره‌ای، رادار ^۳ ، لیدار ^۴ و غیره	داده‌های حرکتی شهروندان از جمله LBSN ^۵
هدف	تشخیص ساختار فیزیکی سطح زمین	کاربری اجتماعی، اقتصادی و غیره
روش‌های پردازش	تصحیح، ادغام، طبقه‌بندی و غیره	پیش پردازش و خوشه‌بندی
شاخص طبقه‌بندی	طیف الکترومغناطیسی و درجه خاکستری	تغییرات زمانی

۲-۲- افزایش محیط‌های شهری در سنجش اجتماعی

به منظور شناخت محیط‌های شهری و سنجش اجتماعی، در مرحله اول باید محیط شهری به بخش‌هایی افزایش شود تا مطالعات بعدی بر روی هر کدام از این بخش‌ها انجام شود [۱۲]. به منظور افزایش محیط‌های شهری در سنجش اجتماعی می‌توان از روش‌های مختلفی استفاده کرد. مثلاً می‌توان از شبکه اصلی راه‌های منطقه شهری به‌عنوان مرز بخش‌های مختلف استفاده کرد [۱۳]. همچنین می‌توان محیط شهری را با یک شبکه منظم سلولی پوشش داد و از این سلول‌ها به‌عنوان واحد مکانی در سنجش اجتماعی بهره گرفت [۹]. در روشی دیگر که

صورت گرفته است. می‌توان به تحقیق نوآلاس و همکاران در سال ۲۰۱۱ اشاره کرد که با استفاده از داده‌های شبکه فوراسکوئر برای دو شهر لندن و نیویورک کاربری اجتماعی را تعیین کردند. در این تحقیق محیط شهری با یک شبکه سلولی منظم پوشش داده شده و از داده‌های معنایی موجود در داده‌های این شبکه بهره‌گیری شد [۹]. همچنین در سال ۲۰۱۳، هاسان و همکاران با استفاده از داده‌های شبکه توییتر و با بهره‌گیری از بعد معنایی داده‌ها به تعیین کاربری محیط شهری پرداختند. آن‌ها با استفاده از تابع چگالی کرنل و برای شش کاربری مختلف و در بازه‌های سه ساعته شبانه‌روز، چگالی هر کاربری را در سطح محیط شهری برآورد کردند [۱۰]. واکامیا و همکاران نیز در سال ۲۰۱۱ پس از افزایش محیط شهری با استفاده از دیاگرام ورونوی بر اساس تعداد داده و تغییرات تعداد داده در طول شبانه‌روز، یک کاربری اجتماعی به هر منطقه منتسب کردند [۲].

در ادامه این مقاله، در بخش دوم مبانی نظری مورد نیاز معرفی می‌گردد. بخش سوم به تشریح داده‌های مورد استفاده می‌پردازد. در بخش چهارم به بیان روش اجرایی انجام تحقیق و ارائه مدل پیشنهادی برای تعیین کاربری‌ها در محیط‌های شهری می‌پردازیم. در بخش پنجم، نتایج حاصل از پیاده‌سازی روش پیشنهادی برای داده‌های کاربران شبکه اجتماعی فوراسکوئر در بخش منهن از شهر نیویورک ارائه شده و مورد ارزیابی قرار خواهد گرفت. درنهایت، در بخش ششم نتیجه‌گیری حاصل از تحقیق بیان شده و پیشنهادهایی برای تحقیقات آینده ارائه خواهد شد.

۲- مبانی نظری تحقیق

در این بخش به تشریح مبانی نظری مورد نیاز در این تحقیق خواهیم پرداخت.

۲-۱- سنجش اجتماعی و کاربری اجتماعی

درک اجتماعی و تعیین کاربری اجتماعی می‌تواند از منابع حرکتی شهروندان حاصل شود. ازجمله این منابع می‌توان به داده‌های مربوط به محل سوار و پیاده شدن از تاکسی، داده‌های حمل‌ونقل عمومی، داده‌های بلیت‌های مترو، داده‌های GPS وسایل نقلیه شخصی و همچنین داده‌های شبکه‌های اجتماعی مکان‌مبنا که مورد توجه این

^۱ Social sensing

^۲ Social land use

^۳ RADAR

^۴ LIDAR

^۵ Location-Based Social Networks

در این تحقیق نیز استفاده شده است، ابتدا داده‌های مکانی کاربران به تعداد معینی خوشه تقسیم می‌شوند و سپس با رسم دیاگرام ورونوی بر روی مراکز این خوشه‌ها محیط شهری به بخش‌هایی افراز خواهد شد و از هر چندضلعی ورونوی می‌توان به‌عنوان واحد مکانی در سنجش اجتماعی بهره گرفت [۱۲].

در این تحقیق به منظور خوشه‌بندی داده‌های مکانی، از شبکه‌های عصبی مصنوعی^۱ استفاده شده است. شبکه‌های عصبی مصنوعی روش‌های محاسباتی نوینی برای یادگیری ماشینی، نمایش دانش و در انتها اعمال دانش به‌دست‌آمده در جهت پیش‌بینی پاسخ‌های خروجی از سامانه‌های پیچیده هستند [۱۴]. یکی از کاربردهای شبکه عصبی، در حوزه یادگیری بدون نظارت و در بحث خوشه‌بندی است، در این راستا از روش یادگیری رقابتی استفاده می‌شود. این شبکه‌ها از تعداد مشخصی نورون تشکیل شده‌اند و هر نورون دارای یک بردار وزن n بعدی است که در آن n برابر با بعد بردارهای ورودی است. بردارهای وزن یا همان سیناپس^۲ ها لایه ورودی را به لایه خروجی (نقشه) متصل می‌کنند. عملکرد شبکه‌های رقابتی در دو فاز رقابت و پاداش صورت می‌گیرد. در فاز رقابت با وارد شدن هر ورودی به شبکه میزان شباهت بین ورودی و نورون‌ها اندازه‌گیری می‌شود (فاصله اقلیدسی در داده‌های مکانی می‌تواند به‌عنوان تابع شباهت به کار رود) سپس مشابه‌ترین نورون به ورودی به‌عنوان نورون برنده انتخاب می‌شود. در فاز پاداش نورون برنده با یک ضریب یادگیری معین که به‌طور کاهشی تغییر می‌کند، پاداش می‌گیرد یعنی به مقدار معینی به ورودی نزدیک می‌شود [۱۴].

یکی از پرکاربردترین روش‌های خوشه‌بندی مکانی مبتنی بر شبکه‌های رقابتی، شبکه رقابتی خود سازمانده^۳ (SOM) است. در این شبکه نورون‌های عصبی در یک ساختار گرید منظم با ابعاد $p \times q$ قرار دارند. تفاوت این شبکه با سایر شبکه‌های رقابتی، اتصال نورون‌ها توسط یک تابع همسایگی است و در فاز پاداش علاوه بر نورون برنده همسایه‌های آن نیز به‌روز می‌شوند. همین اتصال نورون سبب حفظ روابط توپولوژیک بین نورون‌ها شده و یک نقشه خود سازمانده را تولید می‌کند. میزان پاداشی که هر نورون

دریافت می‌کند متأثر از میزان نزدیکی (شباهت) آن نورون به نورون برنده خواهد بود. در رابطه (۱) معادله فاز پاداش در شبکه عصبی خود سازمانده ارائه شده است [۱۵]:

$$w_{t+1}(i) = w_t(i) + \alpha_t h(i, k)(x - w_t(i)) \quad (1)$$

که در آن w_t ، w_{t+1} و i و k به ترتیب وزن نورون در تکرار جدید، وزن نورون در تکرار قبلی، اندکس نورون پاداش گیرنده و اندکس نورون برنده هستند. x ورودی شبکه و α ضریب یادگیری شبکه است که در هر تکرار معمولاً به شکل خطی تغییر می‌کند. همچنین $h(i, k)$ مقدار تابع همسایگی نورون مورد نظر و نورون برنده است و یکی از پرکاربردترین توابع همسایگی که در این مقاله نیز مورد استفاده قرار گرفته است در رابطه (۲) ارائه شده است:

$$h(i, k) = \exp\left(\frac{-\|w(i) - w(k)\|}{2\sigma^2}\right) \quad (2)$$

۲-۳- ارزیابی خوشه‌بندی

در حالت کلی یک خوشه‌بندی بهینه بایستی دارای دو ویژگی فشردگی^۴ (پیوستگی) و تفکیک^۵ باشد. ویژگی فشردگی به معنای میزان شباهت و نزدیکی اعضای متعلق به یک خوشه است و خوشه‌بندی بهینه با حداکثر شدن فشردگی محقق می‌شود؛ و ویژگی تفکیک به معنای میزان شباهت بین خوشه‌های متفاوت است و هر چه این شباهت کمتر باشد خوشه‌بندی مطلوب‌تر خواهد بود [۱۶]. شاخص DB^۶ به‌عنوان یکی از معروف‌ترین و پرکاربردترین شاخص‌های خوشه‌بندی، سعی می‌کند به‌طور هم‌زمان، دو عامل فشردگی و تفکیک را بهینه کند. این شاخص طبق رابطه (۳) محاسبه می‌شود:

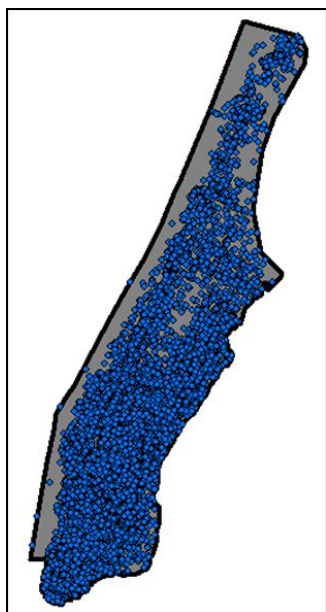
$$DB = \frac{1}{k} \sum_{i=1}^k R_i \quad (3)$$

که در آن k تعداد خوشه‌ها است و R_i از رابطه (۴) محاسبه می‌شود:

$$R_i = \max_{j, j \neq i} \left(\frac{S_i + S_j}{d_{i,j}} \right) \quad (4)$$

^۴ Cohesion
^۵ Separation
^۶ Davies-Bouldin

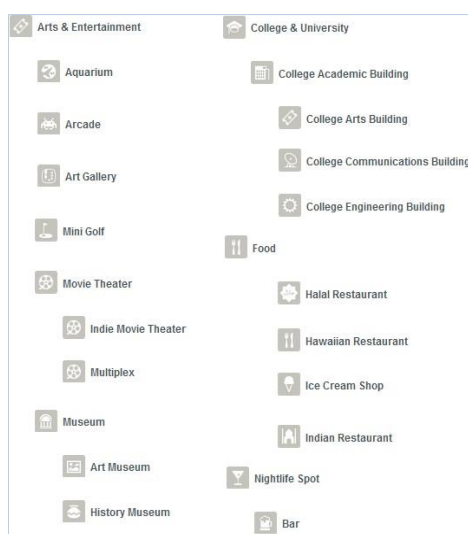
^۱ Artificial Neural Network
^۲ Synapse
^۳ Self-Organization Map



شکل ۱- توزیع مکانی داده‌های مورد مطالعه

۳-۲- بعد معنایی داده‌های شبکه اجتماعی فوراسکوئر

شبکه اجتماعی فوراسکوئر، مکان‌های موجود در محیط شهری را در ۴۰۰ گروه طبقه‌بندی کرده است. این ۴۰۰ گروه به شکل سلسله مراتبی در زیرگروه ۹ بخش اصلی قرار دارند که عبارت‌اند از: ۱- هنر و سرگرمی ۲- دانشگاه‌ها و مراکز آموزشی ۳- غذا ۴- تفریحات ۵- مراکز تفریحی شبانه ۶- اداره‌ها ۷- مسکونی ۸- فروشگاه و خدمات (تجاری) ۹- حمل‌ونقل و سفر. در شکل (۲) بخشی از این طبقه‌بندی نشان داده شده است. هر کاربر زمانی که می‌خواهد از این شبکه استفاده کند بایستی یکی از گروه‌های فوق را انتخاب کند. لذا با بهره‌گیری از این داده‌ها می‌توان کاربری اجتماعی را در مناطق شهری تشخیص داد.



شکل ۲- بخشی از گروه‌بندی مکان‌ها در شبکه اجتماعی فوراسکوئر

در این رابطه، S_i برابر است با میانگین فاصله همه اعضای متعلق به کلاس i از مرکز خوشه و $d_{i,j}$ برابر با فاصله بین مراکز خوشه‌های i و j است. به ازای هر خوشه، $k-1$ مقدار برای R به دست می‌آید که R_i برابر با مقدار ماکزیمم این مقادیر است [۱۷].

۳- داده‌های مورد مطالعه

در این بخش داده‌های مورد مطالعه در دو قسمت معرفی خواهند شد. ابتدا منطقه شهری، شبکه اجتماعی انتخاب شده و داده‌های مکانی آن ارائه می‌شوند و سپس به بررسی بعد معنایی موجود در داده‌های این شبکه خواهیم پرداخت.

۳-۱- منطقه مورد مطالعه و داده‌های مورد استفاده

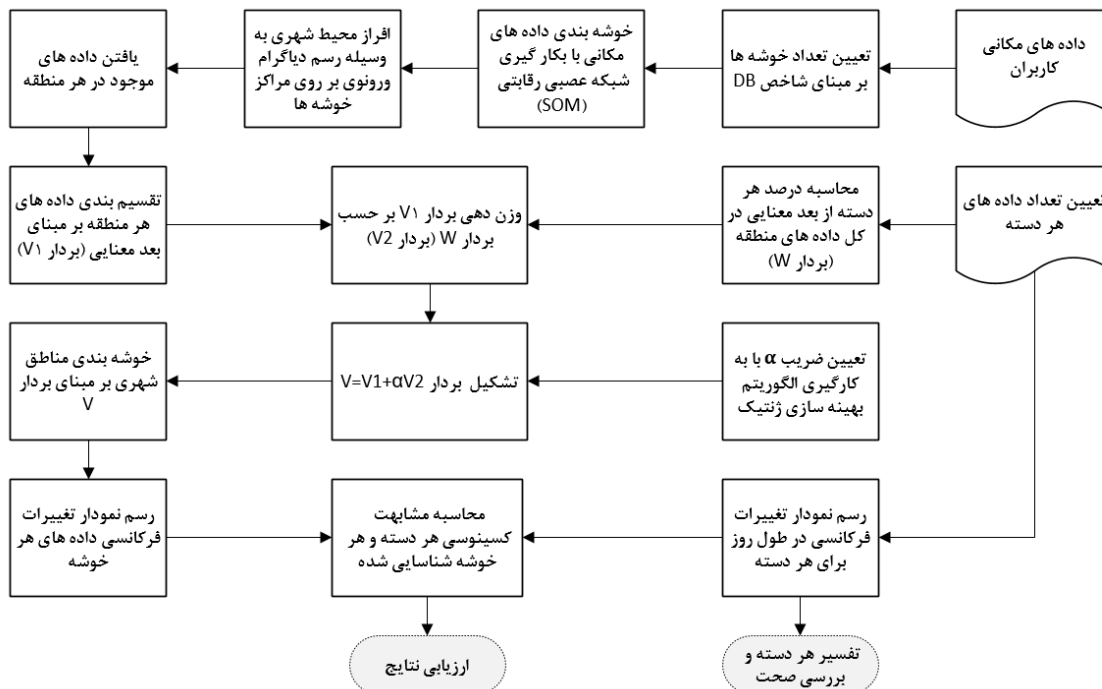
در این تحقیق، بخش منهتن که یکی از پنج بخش شهر نیویورک است، به عنوان محدوده مورد مطالعه انتخاب شده است. جمعیت بخش منهتن در ماه جولای سال ۲۰۱۴، طبق برآورد اداره آمار ایالات متحده حدود ۱/۶۳۶ میلیون نفر و مساحت آن حدود ۸۷/۴۶ کیلومترمربع است. همچنین داده‌های نقطه-مکان^۱ مورد استفاده در این تحقیق از شبکه اجتماعی فراسکوئر جمع‌آوری شده‌اند. این شبکه یکی از معروف‌ترین شبکه‌های اجتماعی مکان‌مبنا است که ۳۱ درصد از کاربران تلفن همراه فعال در شبکه‌های اجتماعی از آن استفاده می‌کنند. این سایت تا ماه اوت سال ۲۰۱۵ بیش از ۵۰ میلیون کاربر فعال ماهانه داشته و حدود ۷ بیلیون داده‌ی نقطه-مکان در آن به اشتراک گذاشته شده است. داده‌های این شبکه اجتماعی علاوه بر ابعاد مکانی و زمانی، شامل بعد معنایی نیز می‌باشد. در این تحقیق، داده‌های کاربران شهر نیویورک در طی حدوداً ۱۰ ماه جمع‌آوری شده و در مرحله بعد داده‌های مربوط به محدوده مورد مطالعه از داده‌های اصلی استخراج شده است، که نتیجه‌ی آن ۱۰۹۳۷۴ داده‌ی نقطه-مکان است. شکل (۱) توزیع مکانی این داده‌ها را در منطقه مورد مطالعه نشان می‌دهد.

^۱ Check-in

۴- روش اجرای تحقیق

معنایی داده‌های کاربران در شبکه‌های اجتماعی مکان‌مبنا، در فلوچارت شکل (۳) نشان داده شده است و در ادامه تشریح خواهد شد.

روش پیشنهادی این تحقیق برای تخصیص کاربری اجتماعی در محیط‌های شهری بر مبنای خوشه‌بندی بعد



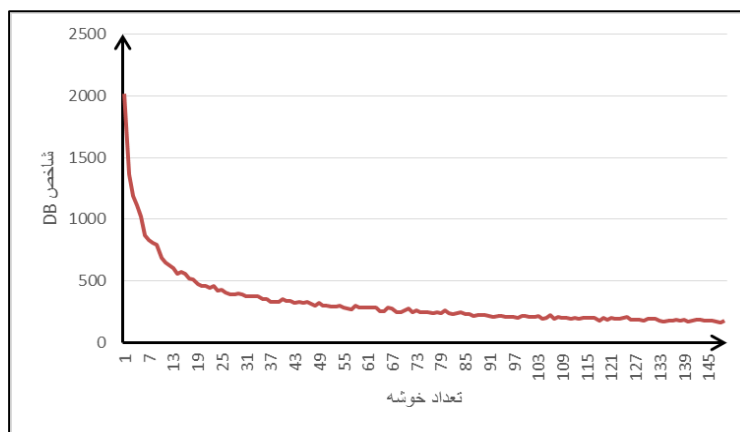
شکل ۳- فلوچارت تعیین کاربری اجتماعی محیط‌های شهری بر مبنای داده‌های شبکه اجتماعی فراسکوئر

۴-۱- تعیین تعداد خوشه‌ها و خوشه‌بندی داده‌های مکانی

بنابراین بهینه‌ترین خوشه‌بندی، دارای کمترین مقدار برای شاخص DB خواهد بود.

در این راستا به ازای $k = 2, 3, \dots, 20$ ، ۲۵ بار خوشه‌بندی K-means تکرار شده و میانگین شاخص DB محاسبه گردید. سپس تعداد خوشه با کمینه به عنوان خوشه‌بندی بهینه انتخاب گردید. نتایج حاصل از اعمال الگوریتم بالا در نمودار شکل (۴) نشان داده شده است.

به منظور خوشه‌بندی داده‌های مکانی کاربران، ابتدا بایستی تعداد بهینه خوشه‌ها محاسبه شود. به این منظور از شاخص DB استفاده گردید. هر چه مقدار این شاخص کاهش پیدا کند، به معنای افزایش فاصله بین خوشه‌های مختلف و کاهش فاصله بین اعضای درون هر خوشه است،



شکل ۴- تغییرات شاخص DB به ازای تعداد خوشه‌ها

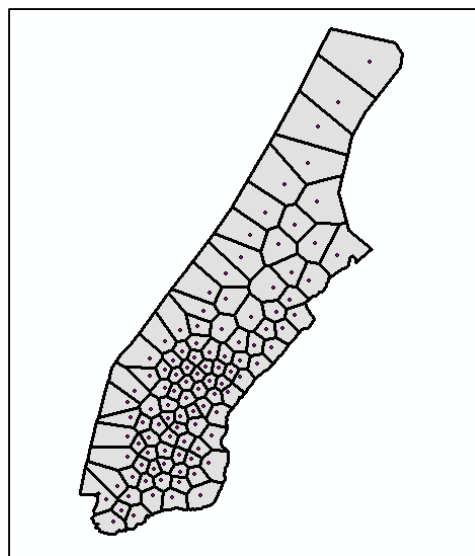
داده‌های معنایی در هر منطقه است؛ اما این شاخص نمی‌تواند به‌خوبی کاربری یک منطقه را تعیین کند، چرا که ممکن است کاربران از یک فعالیت خاص مثلاً تفریح، اطلاعات بیشتری به اشتراک بگذارند تا فعالیتی مانند مسکونی. برای مثال منطقه‌ای را در نظر بگیرید که ۴۰ درصد از داده‌های آن مربوط به گروه غذا و ۳۰ درصد آن مربوط به گروه دانشگاه باشد، بدون شک احتمال تعلق این منطقه به گروه دانشگاهی بیشتر است درحالی‌که ممکن است درصد گروه دیگری بالاتر باشد. به منظور حل این مشکل از معکوس بردار ارائه شده در جدول (۱) به‌عنوان بردار وزن برای اطلاعات هر منطقه استفاده شد (بردار W). حال با محاسبه بردار $V_2 = V_1 \times W$ می‌توان در واقع از نسبت فراوانی هر گروه معنایی در یک منطقه خاص به فراوانی آن گروه در کل داده‌ها به‌عنوان شاخص جدید کاربری اجتماعی بهره گرفت؛ اما این بردار هم نمی‌تواند به‌خوبی نمایان گر کاربری یک منطقه باشد. برای روشن شدن موضوع منطقه‌ای را در نظر بگیرید که ۷ درصد داده‌های آن متعلق به گروه معنایی دانشگاه بوده و ۶۰ درصد آن متعلق به گروه غذا باشد، با اعمال ماتریس وزن، اعداد تقریبی ۳ و ۲ به ترتیب برای گروه دانشگاه و غذا حاصل می‌شود اما نمی‌توان گفت که این منطقه متعلق به گروه دانشگاهی است. از طرف دیگر ممکن است هر کدام از دو شاخص معرفی شده در شناسایی گروه‌های معنایی در ۹ خوشه نهایی ناتوان باشند، یعنی یک گروه معنایی دو و یا چند بار به‌عنوان مؤلفه حداکثری مرکز خوشه شناسایی شود و در مقابل، گروه معنایی وجود داشته باشد که هرگز شناسایی نشود.

۴-۳- معرفی شاخص خوشه‌بندی

ایده اصلی این تحقیق، خوشه‌بندی مناطق شناسایی شده بر مبنای داده‌های معنایی است. در این راستا، معرفی شاخص مناسب برای خوشه‌بندی از اهمیت بسزایی برخوردار است. تعداد خوشه‌ها از قبل مشخص و برابر ۹ است. توجه به این نکته بسیار ضروری است که الگوریتم خوشه‌بندی باید مراکز خوشه‌ای تولید کند که در هر خوشه مؤلفه‌ی مربوط به یک گروه معنایی منحصربه‌فرد حداکثر شود؛ یعنی هیچ کاربری معنایی نباید دو بار تشخیص داده شود درحالی‌که گروه دیگری هرگز شناسایی نشود؛ بنابراین باید شاخصی برای خوشه‌بندی معرفی شود

همان‌طور که در نمودار شکل (۴) مشاهده می‌شود، مقدار شاخص DB از حدود تکرار صدم به یک مقدار ثابت می‌رسد؛ بنابراین باید تعداد نوروهای شبکه رقابتی SOM نزدیک به عدد ۱۰۰ باشد. حال با اختیار کردن اعداد مختلف برای پارامترهای p و q در شبکه رقابتی SOM و محاسبه شاخص DB بهترین مقادیر برای تعداد نوروها شبکه $p=15$ و $q=7$ به دست آمدند. در نهایت با به‌کارگیری الگوریتم خوشه‌بندی SOM داده‌های مکانی به ازای $k = p \times q = 105$ خوشه‌بندی می‌شوند.

پس از اعمال الگوریتم خوشه‌بندی می‌توان از مراکز خوشه به‌عنوان نقاط نمونه استفاده کرد و با ترسیم دیاگرام ورونوی برای این نقاط منطقه مد نظر را به چند زیر منطقه افراز کرده و در مراحل بعدی از چندضلعی‌های این دیاگرام به عنوان واحد مکانی برای مطالعه منطقه شهری استفاده کرد. در شکل (۵) مراکز خوشه بندی و دیاگرام ورونوی متناظر با آن‌ها نشان داده شده است.



شکل ۵- مناطق شناسایی شده در محیط شهری بر مبنای رسم دیاگرام ورونوی مراکز خوشه‌های مکانی

۴-۲- بررسی نحوه توزیع گروه‌های معنایی در مناطق شهری و تأثیر آن در کاربری اجتماعی

در این مرحله تعداد داده‌های موجود در هر منطقه به تفکیک گروه‌های معنایی محاسبه شد و سپس بردار مربوط به هر منطقه نرمال گردید (بردار V_1). مشابه همین عملیات برای کل داده‌ها در تمام محیط شهری مورد مطالعه تکرار شد که نتایج آن در جدول (۲) ارائه شده است. ایده اولیه، استفاده از فراوانی هر گروه از

که اولاً هر ۹ گروه معنایی را شناسایی کرده و ثانیاً مؤلفه مربوط به خوشه شناسایی شده را حداکثر کند. به منظور روشن تر شدن این موضوع فرض کنید الگوریتم خوشه‌بندی یک ماتریس 9×9 برای مراکز خوشه ایجاد کرده است که در آن خوشه k ام گروه معنایی k ام را شناسایی کرده باشد، حالت ایده آل خوشه‌بندی زمانی اتفاق می‌افتد که عناصر قطر اصلی برابر ۱ و سایر درایه‌ها صفر باشند. بنابراین به منظور بهینه‌سازی عملیات خوشه‌بندی به دنبال شاخصی برای خوشه‌بندی خواهیم بود که در آن مجموع مؤلفه‌های بیشینه (مربوط به گروه معنایی شناسایی شده) در هر مرکز خوشه حداکثر شود و ۸ مؤلفه باقی‌مانده تا حد امکان یکنواخت بوده و فاقد پراکندگی باشند. مثلاً مرکز خوشه با مؤلفه‌های (0.6 0.05 0.05 0.05 0.05 0.05 0.05 0.05 0.05) نسبت به مرکز خوشه دیگری با مؤلفه‌های (0.6 0 0 0 0.4 0 0 0 0) مطلوب‌تر است. لذا در این تحقیق شاخص $V = V_1 + \alpha V_2$ به شکل ترکیب خطی از دو شاخص V_1 و V_2 که پیش از این معرفی شدند، به نحوی به منظور خوشه‌بندی به کار گرفته شد که ویژگی‌های مذکور را به بهترین نحو برآورده کند.

جدول ۲- توزیع داده معنایی در کل منطقه

گروه معنایی	درصد فراوانی
هنر و سرگرمی	۵/۹
دانشگاه و مراکز آموزشی	۲/۶
غذا	۲۷/۶
تفریح	۷/۶
مراکز تفریحی شبانه	۱۰
اداره‌ها	۱۴/۵
مسکونی	۳/۵
فروشگاه و خدمات (تجاری)	۱۸
حمل‌ونقل و سفر	۱۰/۳

۴-۴- به‌کارگیری الگوریتم بهینه‌سازی ژنتیک برای یافتن مقدار بهینه برای ضریب α

در این تحقیق به منظور یافتن مقدار بهینه α در ترکیب خطی $V = V_1 + \alpha V_2$ از الگوریتم ژنتیک بهره گرفته شده است. در نسل اول ۵۰ مقدار تصادفی بین ۰ تا ۱۰۰ برای این ضریب تولید می‌شود؛ در هر تکرار بهترین

جواب به نسل بعد منتقل می‌شود و سایر اعضای جمعیت نسل بعدی با عملیات ترکیب و جهش تولید می‌شوند. هر یک از عناصر الگوریتم به کار رفته در ادامه توضیح داده می‌شوند [۱۸]:

تابع برازندگی: به ازای هر مقدار α شاخص V محاسبه می‌شود و عملیات خوشه‌بندی به روش k-means بر روی مناطق شهری اعمال می‌شود. اگر مرکز خوشه تولید شده همه گروه‌های معنایی را شناسایی نماید، مقدار تابع برازندگی آن از رابطه (۵) محاسبه می‌شود در غیر این صورت آن گزینه حذف خواهد شد:

$$fit = \frac{\sum_{i=1}^9 \max(C_i)}{\sigma_C} \quad (5)$$

که در آن C مرکز خوشه و C' مرکز خوشه به‌جز مؤلفه ماکزیمم است. آن چنانکه پیش‌تر بیان شد هر چه مقدار مؤلفه حداکثری خوشه بیشتر باشد و انحراف معیار سایر مؤلفه‌ها کمتر باشد، خوشه‌بندی مطلوب‌تر خواهد بود.

انتخاب والدین و ترکیب: به این منظور از الگوریتم چرخ رولت استفاده می‌شود و ۹ عضو از جمعیت نسل بعد با گرفتن میانگین وزن‌دار از والدین انتخاب شده تولید می‌شوند.

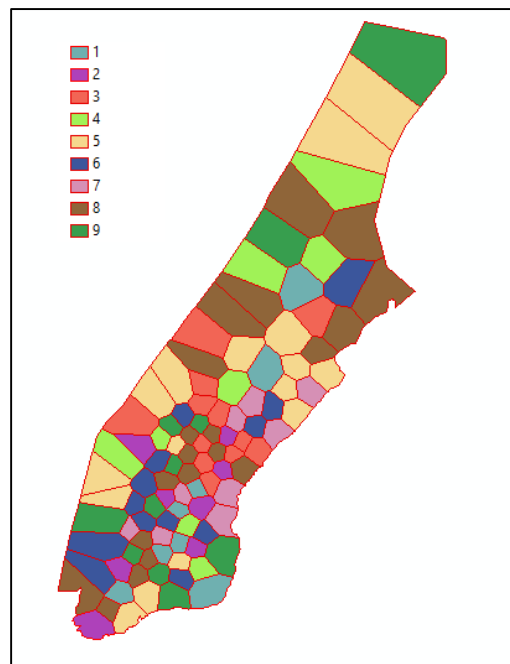
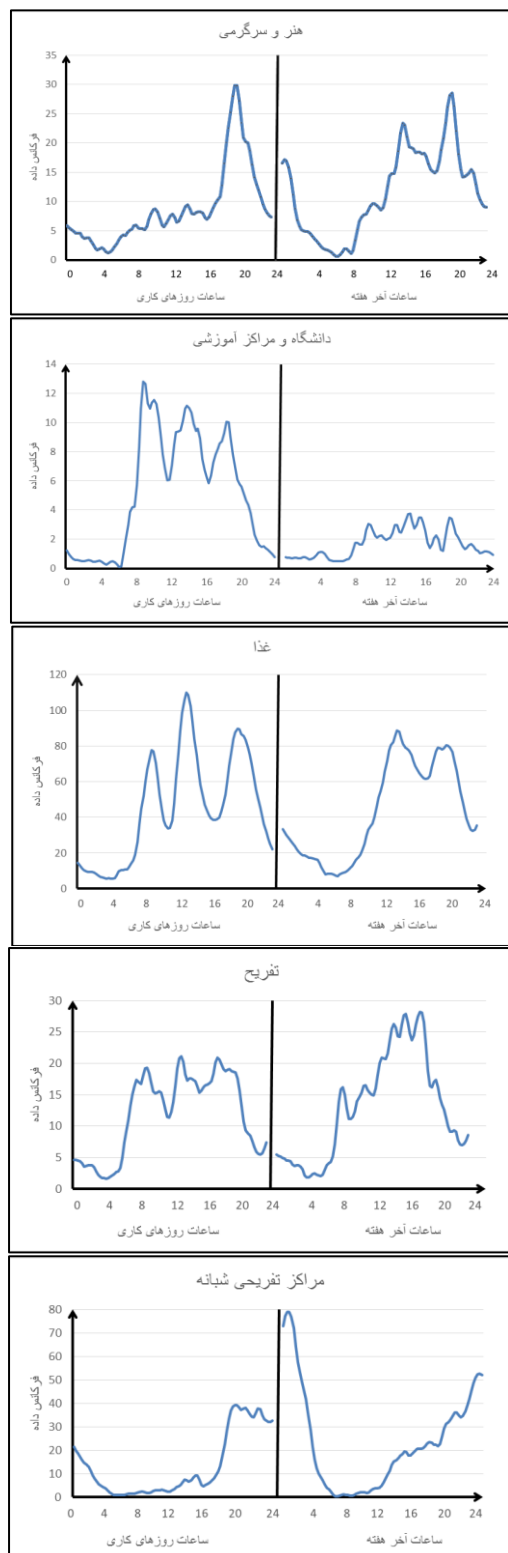
جهش: ۴۰ عضو باقی‌مانده نیز مجدداً به‌طور اتفاقی تولید می‌شوند.

در نهایت مقدار $4/27$ برای ضریب α تعیین شد.

۴-۵- خوشه‌بندی مناطق شهری بر مبنای شاخص تعیین شده

پس از یافتن شاخص مناسب خوشه‌بندی، خوشه‌بندی به روش k-means به تعداد ۱۰۰ بار برای مناطق شهری تکرار شد و خوشه‌بندی ای که معیار معرفی شده در رابطه (۵) را بیشینه نمود به‌عنوان خوشه‌بندی نهایی انتخاب شد. در شکل (۶) نقشه منطقه مورد مطالعه به تفکیک مناطق شناسایی شده نشان داده شده است؛ هم‌چنین نمودار تغییرات فرکانسی مربوط به هر یک از خوشه‌های شناسایی شده در شکل (۷) نشان داده شده است.

داده‌های مربوط به هر یک از خوشه‌های شناسایی شده در شکل (۸) نشان داده شده است. در ادامه به تحلیل نمودارهای شکل (۷) می‌پردازیم:



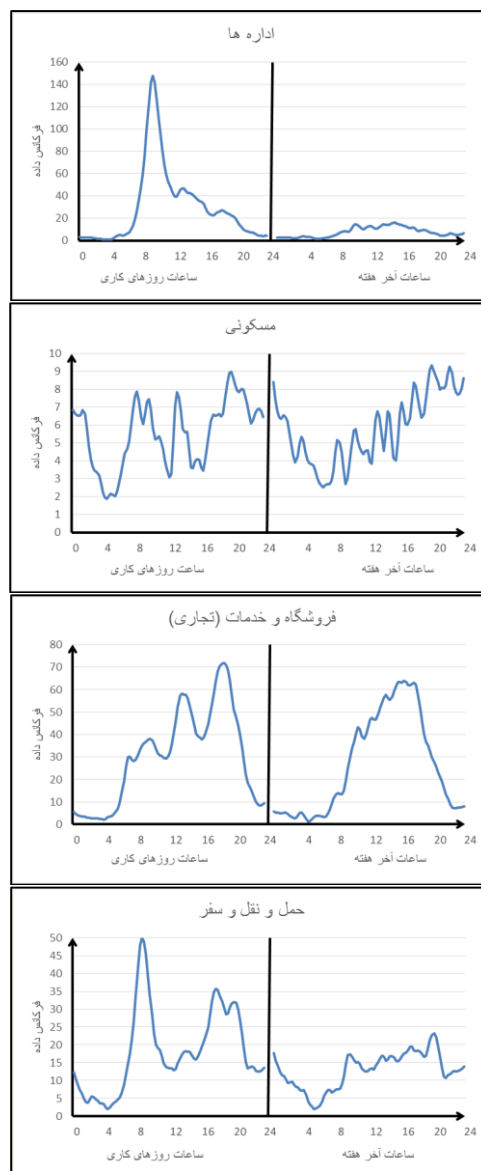
شکل ۶- مناطق شناسایی شده در محیط شهری بر مبنای رسم دیاگرام ورونوی مراکز خوشه‌های مکانی

۵- ارزیابی

در روش پیشنهادی این تحقیق ابتدا محیط شهری به چند زیر بخش افراز گردید و سپس بر مبنای شاخص معرفی شده به هر بخش یک کاربری اجتماعی اختصاص داده شد. در ادامه صحت داده‌هایی که کاربران به اشتراک گذاشته‌اند بررسی می‌گردد، هم‌چنین ضریب همبستگی بین نمودار تغییرات زمانی دسته‌های معنایی و نمودار تغییرات زمانی خوشه‌های شناسایی شده به‌عنوان شاخص ارزیابی این تحقیق معرفی خواهد شد.

۵-۱- داده‌کاوی روی داده‌های معنایی و ارزیابی صحت

در این مرحله به‌منظور بررسی صحت داده‌هایی که کاربران از نوع فعالیت خود در شبکه اجتماعی فوراسکوئر به اشتراک می‌گذارند، تعداد داده‌های هر دسته به ازای هر ۱۵ دقیقه در طول شبانه‌روز محاسبه شد. به منظور بررسی دقیق‌تر، روزهای کاری هفته و روزهای پایان هفته (شنبه و یکشنبه) به شکل جداگانه مورد بررسی قرار گرفتند؛ یعنی به ازای هر دسته یک بردار ۱۹۲ عضوی تشکیل شد. شکل (۷) نمودار تغییرات فرکانسی برای ۹ گروه معرفی شده را نمایش می‌دهد. هم‌چنین نمودار تغییرات فرکانسی



شکل ۷- نمودار تغییرات فرکانسی هر گروه از داده‌های شبکه اجتماعی فراسکوئر در طول روز

گروه ۱- هنر و سرگرمی: برای روزهای کاری هفته تعداد این گروه از داده‌ها، طبق انتظار، پس از ساعت ۲۴:۰۰ کاهش پیدا می‌کند و این روند تا ساعت ۵:۰۰ ادامه دارد (زمان خواب)، سپس روند افزایشی خود را شروع کرده و از ساعت ۱۷:۰۰ با شیب زیادی به یک مقدار ماکزیمم در ساعت ۲۰:۳۰ می‌رسد و بار دیگر روند کاهشی آغاز می‌شود و تا ساعت ۵:۰۰ فردا ادامه پیدا می‌کند. همان‌طور که انتظار می‌رفت تعداد داده‌های سرگرمی در ساعات عصر به شکل چشم‌گیری افزایش پیدا کرده است. در روزهای پایان هفته نیز تعداد داده‌ها از ساعت ۲۴:۰۰ روند کاهشی خود را آغاز می‌کند اما این روند تا ساعت ۸:۰۰ ادامه دارد. همچنین فراوانی داده‌های این گروه در پایان هفته به شکل قابل‌توجهی

بیشتر از روزهای کاری هفته است و در ادامه برخلاف روزهای کاری هفته قبل از ساعت ۱۷:۰۰ نیز شاهد یک بیشینه محلی هستیم و در ادامه روز، الگوی نسبتاً مشابه با روزهای کاری هفته تکرار می‌شود.

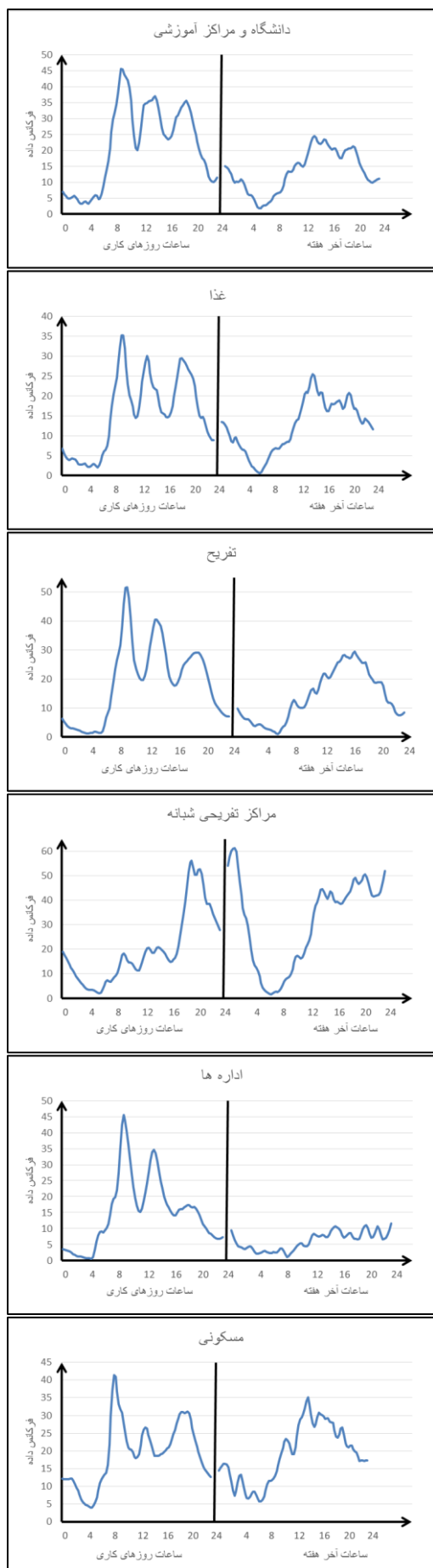
گروه ۲- دانشگاه و مراکز آموزشی: فراوانی داده‌های این گروه از ساعت ۲۴:۰۰ الی ۷:۰۰ بسیار ناچیز است و پس از آن به شکل ناگهانی افزایش پیدا کرده و دارای سه بیشینه محلی است که احتمالاً مربوط به زمان استراحت و وعده‌های غذایی هستند. فراوانی این داده‌ها از ساعت ۱۸:۰۰ روند کاهش خود را آغاز می‌کند. در روزهای آخر هفته نیز مطابق انتظار تعداد داده‌ها بسیار ناچیز است و همان مقدار کم نیز الگوی نسبتاً مشابه روزهای کاری هفته دارند.

گروه ۳- غذا: به شکل کاملاً واضحی نمودار این گروه در روزهای کاری هفته سه بیشینه محلی در ساعات ۹:۳۰، ۱۳ و ۱۹:۳۰ دارد که منطبق بر وعده‌های غذایی هستند. همچنین در روزهای آخر هفته مطابق انتظار وعده صبحانه از نمودار حذف شده و دو بیشینه محلی دیگر کمی به جلوتر و به ساعات‌های ۱۴:۰۰ و ۲۰:۰۰ انتقال یافته‌اند.

گروه ۴- تفریح: با مقایسه روزهای کاری و روزهای آخر هفته برای این گروه اولاً تعداد بیشتر داده در روزهای آخر هفته و ثانیاً تأخیر در شروع افزایش صبحگاهی در روزهای آخر هفته جلب توجه می‌کند. همچنین افزایش چشم‌گیر تعداد داده‌های این گروه در عصر روزهای تعطیل به درستی این اطلاعات صحنه می‌گذارد.

گروه ۵- مراکز تفریحی شبانه: در روزهای کاری هفته مطابق انتظار، نمودار با روند کاهشی از ساعت ۲۴:۰۰ آغاز می‌شود و به مقدار کمی (نزدیک به صفر) می‌رسد. نمودار با همین فراوانی اندک ادامه پیدا می‌کند و ناگهان از ساعت ۱۷ روند افزایشی خود را آغاز کرده و به یک مقدار تقریباً ثابت از ساعت ۲۰:۰۰ الی ۲۴:۰۰ می‌رسد؛ اما در روزهای پایان هفته روند افزایشی از ساعت ۱۲:۰۰ آغاز می‌شود و تا ساعت ۱:۰۰ بامداد ادامه پیدا می‌کند. فراوانی چشم‌گیر این داده‌ها در پایان هفته نسبت به روزهای کاری از نکات قابل توجه و مهم این نمودار است.

گروه ۶- اداره‌ها: مطابق انتظار در ساعات بامداد فراوانی این داده‌ها نزدیک به صفر است و از ساعت ۵:۰۰ روند افزایش ناگهانی خود را آغاز می‌کند و در ساعت ۱۰:۰۰ به مقدار بیشینه خود می‌رسد. پس از آن و با



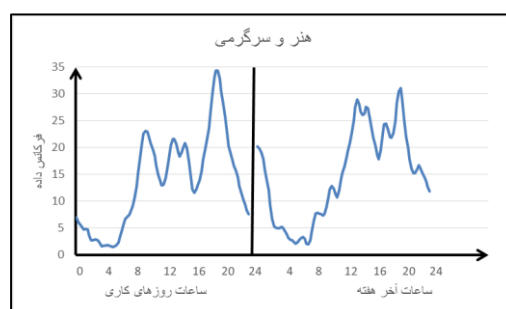
نزدیک شدن به انتهای ساعات کاری روز فراوانی این داده‌ها کاهش پیدا می‌کند. همان‌طور که در شکل (۸) نیز مشخص است، در روزهای تعطیل هفته تعداد این داده‌ها بسیار ناچیز است.

گروه ۷- مسکونی: از نمودار این گروه الگوی چندان منظمی نمی‌توان استنباط کرد اما سه بیشینه محلی در ساعات نزدیک به وعده‌های غذایی و همچنین افزایش فراوانی داده‌ها در ساعات پایانی روز (بر خلاف سایر گروه‌ها)، همراه با بازگشت شهروندان به خانه، از نکات قابل توجه این نمودار است.

گروه ۸- فروشگاه و خدمات (تجاری): مطابق انتظار در روزهای کاری هفته افزایش فعالیت شهروندان در حوزه تجاری از ساعت ۶:۰۰ روند صعودی خود را آغاز می‌کند و پس از یک کمینه محلی برای نهار و استراحت دوباره روند افزایشی پیدا می‌کند. همان‌طور که در شکل (۸) نیز مشخص است برای روزهای آخر هفته روند افزایشی صبحگاهی دیرتر آغاز می‌شود و کمی زودتر در ساعت ۱۸:۰۰ روند کاهشی آغاز می‌گردد.

گروه ۹- حمل‌ونقل و سفر: فراوانی داده‌ها در این گروه از ساعت ۵:۰۰ صبح و با شروع فعالیت‌های تجاری، اداری، آموزشی و غیره شهروندان افزایش پیدا کرده و پس از رسیدن شهروندان به محل کار خود به یک مقدار نسبتاً کمی رسیده و بار دیگر برای بازگشت شهروندان به منازل اوج می‌گیرد. در روزهای پایان هفته نیز مطابق انتظار فراوانی این داده به شکل چشمگیری کم می‌شود و الگوی نسبتاً مشابهی تکرار می‌شود. این نمودار تا حدودی منطبق بر میزان ترافیک شهری مورد انتظار است.

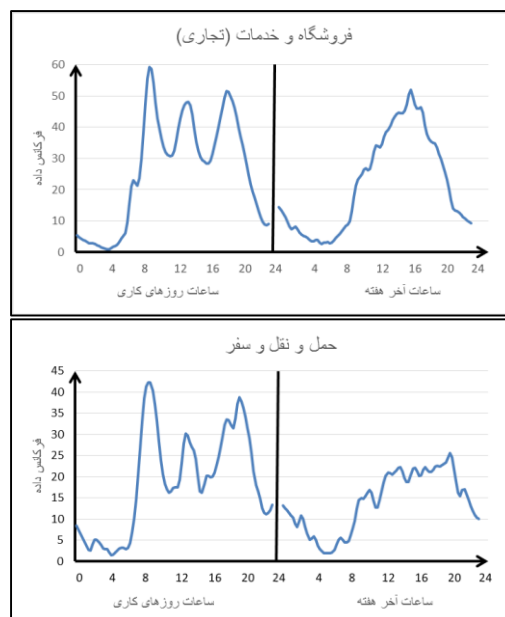
داده‌کاوای صورت گرفته نشان می‌دهد که داده‌های شبکه‌های اجتماعی به شکل قابل قبولی بر الگوی رفتاری شهروندان در گروه‌های معنایی معرفی شده منطبق هستند و می‌توان از این داده‌ها در راستای شناخت محیط‌های شهری سود جست.



۶- نتیجه‌گیری

در این تحقیق، از بعد معنایی داده‌های شهروندان در شبکه‌های اجتماعی به‌عنوان شاخصی برای تعیین نوع کاربری اجتماعی مناطق مختلف شهری استفاده گردید. برای این منظور پس از خوشه‌بندی داده‌های مکانی کاربران با رسم دیاگرام ورونوی بر روی مراکز خوشه‌های شناسایی شده، محیط شهری به چندین زیر بخش افراز گردید، سپس بر مبنای شاخصی معرفی شده بر اساس بعد معنایی داده‌های کاربران، مناطق شهری خوشه‌بندی شده و به هر منطقه یک کاربری اجتماعی نسبت داده شد. برای ارزیابی روش پیشنهادی، نمودار تغییرات زمانی داده‌های کاربران در مناطق شناسایی شده و هم‌چنین گروه‌های معنایی موجود در داده‌ها ترسیم گردید. لازم به ذکر است با توجه به تغییر رفتار اجتماعی شهروندان در روزهای آخر هفته این نمودار در دو بخش روزهای کاری و روزهای آخر هفته ترسیم شد. سپس از ضریب همبستگی بین نمودارهای به منظور ارزیابی روش پیشنهادی استفاده شد. نتایج حاصل از این تحقیق و ارزیابی انجام شده توجه ما را به نکات زیر جلب می‌کند:

- داده‌های انبوه شبکه‌های اجتماعی علاوه بر هزینه کم و به‌روز بودن، به طور صادقانه‌ای توسط شهروندان به اشتراک گذاشته می‌شود و از صحت مناسبی برخوردار است.
- مناطق شهری با کاربری اجتماعی یکسان و یا مشابه دارای پیوستگی مکانی هستند.
- نقشه خروجی این تحقیق مطابقت خوبی با واقعیت موجود در منطقه منهن دارد.
- ارزیابی این تحقیق توانایی شبکه‌های اجتماعی برای شناخت محیط‌های شهری را تأیید می‌کند. برای تحقیقات آینده می‌توان از داده‌های حرکتی سایر منابع نظیر داده‌های GPS خودروها استفاده کرد و نتایج حاصل را با هم مقایسه نمود. استفاده از تقسیمات سیاسی شهرها و هم‌چنین یک شبکه سلولی منظم به‌عنوان واحد مکانی نیز می‌تواند به نتایج قابل توجهی منجر شود. هم‌چنین استفاده از داده‌های شبکه‌های اجتماعی مکان‌مبنا در راستای حل معضلات شهری نظیر ترافیک، آلودگی صوتی، آلودگی هوا و غیره نیز می‌تواند چشم‌انداز جدیدی چه در زمینه پژوهشی و چه در زمینه تجاری باشد.



شکل ۸- نمودار تغییرات فرکانسی داده‌های خوشه‌های شناسایی شده

۵-۲- محاسبه ضریب همبستگی نمودار خوشه‌های شناسایی شده و دسته‌های معنایی

همان‌طور که انتظار می‌رفت نمودار تغییرات زمانی هر گروه معنایی شکل (۷) و نمودار تغییرات زمانی خوشه‌های شناسایی شده شکل (۸) رفتار نسبتاً مشابهی در طول ساعات شبانه‌روز برای روزهای کاری و آخر هفته دارند. به منظور ارزیابی روش پیشنهادی ضریب همبستگی بین نمودارها محاسبه و در جدول (۲) ارائه شده است. میانگین ضرایب به دست آمده ۰/۸۳ است و این همبستگی بالای بین نمودارهای مذکور نشان دهنده رفتار مشابه دو نمودار بوده و این ارزیابی به‌خوبی توانایی بالای شبکه‌های اجتماعی برای سنجش اجتماعی محیط‌های شهری را روشن می‌کند.

جدول ۳- ضریب همبستگی نمودارهای گروه‌های معنایی و خوشه‌های شناسایی شده

گروه معنایی	ضریب همبستگی
هنر و سرگرمی	۰/۸۴۱
دانشگاه و مراکز آموزشی	۰/۸۷۲
غذا	۰/۸۷۵
تفریح	۰/۸۱۰
مراکز تفریحی شبانه	۰/۸۴۵
اداره‌ها	۰/۸۷۳
مسکونی	۰/۵۴۶
فروشگاه و خدمات (تجاری)	۰/۹۲۰
حمل‌ونقل و سفر	۰/۸۸۹

- [1] Y. Zheng, L. Capra, O. Wolfson, and H. Yang, "Urban computing: concepts, methodologies, and applications," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 5, p. 38, 2014.
- [2] S. Wakamiya, R. Lee, and K. Sumiya, "Urban area characterization based on semantics of crowd activities in twitter," in *International Conference on GeoSpatial Semantics*, 2011, pp. 108-123.
- [3] A. Kheiri, F. Karimipour, and M. Forghani, "Intra-Urban Movement Patterns Estimation Based on Location Based Social Networking Data," *Journal of Geomatics Science and Technology*, vol. 6, pp. 141-158, 2016.
- [4] Q. Li, Y. Zheng, X. Xie, Y. Chen, W. Liu, and W.-Y. Ma, "Mining user similarity based on location history," in *Proceedings of the 16th ACM SIGSPATIAL international conference on Advances in geographic information systems*, 2008, p. 34.
- [5] Y. Zheng, Y. Chen, X. Xie, and W.-Y. Ma, "GeoLife2. 0: a location-based social networking service," in *2009 Tenth International Conference on Mobile Data Management: Systems, Services and Middleware*, 2009, pp. 357-358.
- [6] N. Li and G. Chen, "Analysis of a location-based social network," in *Computational Science and Engineering, 2009. CSE'09. International Conference on*, 2009, pp. 263-270.
- [7] S. Scellato, C. Mascolo, M. Musolesi, and V. Latora, "Distance Matters: Geo-social Metrics for Online Social Networks," in *WOSN*, 2010.
- [8] J. Cranshaw, R. Schwartz, J. I. Hong, and N. Sadeh, "The Livehoods Project: Utilizing Social Media to Understand the Dynamics of a City," in *Proceedings of the Sixth International AAAI Conference on Weblogs and Social Media (ICWSM 2012)*, Trinity College in Dublin, Ireland, 2012.
- [9] A. Noulas, S. Scellato, C. Mascolo, and M. Pontil, "Exploiting Semantic Annotations for Clustering Geographic Areas and Users in Location-based Social Networks," in *Fifth International AAAI Conference on Weblogs and Social Media*, 2011.
- [10] S. Hasan, X. Zhan, and S. V. Ukkusuri, "Understanding urban human activity and mobility patterns using large-scale location-based data from online social media," in *Proceedings of the 2nd ACM SIGKDD international workshop on urban computing*, 2013, p. 6.
- [11] Y. Liu, X. Liu, S. Gao, L. Gong, C. Kang, Y. Zhi, et al., "Social sensing: A new approach to understanding our socioeconomic environments," *Annals of the Association of American Geographers*, vol. 105, pp. 512-530, 2015.
- [12] V. Frias-Martinez and E. Frias-Martinez, "Spectral clustering for sensing urban land use using Twitter activity," *Engineering Applications of Artificial Intelligence*, vol. 35, pp. 237-245, 2014.
- [13] N. J. Yuan, Y. Zheng, and X. Xie, "Segmentation of urban areas using road networks," *MSR-TR-2012-65*, Tech. Rep., 2012.
- [14] B. Yegnanarayana, *Artificial neural networks*: PHI Learning Pvt. Ltd., 2009.
- [15] T. Kohonen, "The self-organizing map," *Neurocomputing*, vol. 21, pp. 1-6, 1998.
- [16] S. Das, A. Abraham, and A. Konar, "Automatic clustering using an improved differential evolution algorithm," *IEEE Transactions on systems, man, and cybernetics-Part A: Systems and Humans*, vol. 38, pp. 218-237, 2008.
- [17] D. L. Davies and D. W. Bouldin, "A cluster separation measure," *IEEE transactions on pattern analysis and machine intelligence*, pp. 224-227, 1979.
- [18] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," *IEEE transactions on evolutionary computation*, vol. 6, pp. 182-197, 2002.