بازیابی و تحلیل خصوصیات شهرهای مختلف در شبکه‌های اجتماعی مبتنی بر مکان با استفاده از الگوریتم جستجوی هارمونی

مجید عبدالرزاق نژاد[۱]، حامد وحدت نژاد[۲]، محمد قانعی استاد[۳]

[۱]استادیار گروه مهندسی کامپیوتر، دانشکدهٔ فنی مهندسی، دانشگاه بزرگمهر قائنات، abdolrazzagh@buqaen.ac.ir

[۲]استادیار گروه مهندسی کامپیوتر، دانشکدهٔ برق و کامپیوتر، دانشگاه بیرجند، vahdatnejad@birjand.ac.ir

[۳]دانشجوی کارشناسی ارشد فناوری اطلاعات، دانشکدهٔ برق و کامپیوتر، دانشگاه بیرجند، m.ghanei1992@birjand.ac.ir

چکیده

با افزایش محبوبیت شبکه‌های اجتماعی مبتنی بر مکان[۱] و دسترسی به ردپای دیجیتالی[۲] کاربران، یک فرصت مطالعاتی وسیع برای تحلیل رفتار مردم شهرهای مختلف فراهم شده است. ساختار گروه‌ها در شبکه‌های اجتماعی یکی از مهم‌ترین مشخصات آن‌هاست. برخلاف سایر شبکه‌های اجتماعی که گروه‌ها بطور صریح برای عضویت مشخص هستند، در LBSN ها، گروه‌ها تعریف دقیقی ندارند. با توجه به جذابیت بررسی رفتار گروهی کاربران، کشف انجمن[۳] در LBSN ها، توجه بسیاری از محققان را به خود جلب کرده است. در این مقاله یک چارچوب هوشمند برای کشف انجمن‌های هم‌پوشا[۴] در LBSN ها بر مبنای بازدیدهای انجام شده توسط کاربران و ویژگی‌های[۵] محل‌های بازدید و کاربران، ارائه می‌شود. همچنین خوشه‌بندی بر روی یال‌های دوقسمتی انجام شده که شامل کاربران و محل‌های بازدید هستند. سپس هر انجمن با استفاده از فرادادهٔ[۶] کاربران و محل‌های بازدید، پروفایل شده و در گام بعد مشخصات شهرها استخراج می‌شوند. تمرکز این مقاله بر روی داده‌های به دست آمده از شبکهٔ Foursquare می‌باشد. مزیت رویکرد پیشنهادی، استفاده از الگوریتم فرامکاشفه‌ای[۷] جستجوی هارمونی[۸] با هدف تعیین هوشمند مراکز خوشه‌ها است.

واژه‌های کلیدی

الگوریتم جستجوی هارمونی، انجمن‌های هم‌پوشا، پروفایل انجمن، خصوصیات شهرها، شبکه‌های اجتماعی مبتنی بر مکان

---

[1] . Location Based Social Network (LBSN)

[2] . Digital Footprint

[3] . Community

[4] . Overlapping Communities

[5] . Attribute

[6] . Metadata

[7] . Meta Heuristic Algorithms

[8] . Harmony Search

۱- مقدمه

در سال‌های اخیر، با پیشرفت تکنولوژی موبایل‌ها، شبکه‌های اجتماعی مبتنی بر مکان بسیار محبوب شده و کاربران بسیاری را به خود جذب کرده‌اند [1] [2]. LBSN ها، علاوه بر پشتیبانی از امکانات مختلفی نظیر دوست‌یابی و اشتراک‌گذاری متن و تصویر، تعاملات آنلاین همانند بازدید از یک محل را هم فراهم می‌کنند [3] [4]. کاربران می‌توانند محل‌های مختلف را جستجو نموده و نظرات خود را دربارهٔ آن‌ها بنویسند. بطورکلی LBSN ها یک ساختار اجتماعی جدید هستند که اعضای آن بر اساس مکان‌هایی که در دنیای واقعی بازدید کرده‌اند در ارتباط می‌باشند [3]. با افزایش محبوبیت این شبکه‌ها، فرصت جدیدی برای مطالعه بر روی رفتار گروهی مردم فراهم می‌شود [5].

ساختار گروه‌ها در LBSN ها، بسیار با اهمیت بوده و مورد توجه محققان قرار گرفته است. در شبکه‌های اجتماعی، این گروه‌ها به عنوان «خوشه» یا «انجمن» هم شناخته می‌شوند [6] [7]. بطور کلی یک «انجمن به گروهی از کاربران یک شبکه گفته می‌شود که نسبت به سایر اعضای شبکه، شدیدا در ارتباط هستند» [8]. برخلاف سایر شبکه‌های اجتماعی که گروه‌ها بطور صریح برای عضویت مشخص می‌باشند، در LBSN ها، گروه‌ها بطور واضح تعریف نشدند. از اینرو کشف انجمن در LBSN ها، به یک فرصت ویژه برای شناسایی و تحلیل خصوصیات شهرهای مختلف تبدیل می‌شود.

در بسیاری از روش‌های ارائه شده برای کشف انجمن، ویژگی‌های ساختاری شبکه‌ها مد نظر قرار گرفته شده‌است [9]. اما تنها درنظر گرفتن ویژگی‌های ساختاری، نتایج ضعیفی به همراه دارد، زیرا کاربران در گروه-های جدا از هم قرار می‌گیرند. باتوجه به اینکه در سایر شبکه‌های اجتماعی، کاربران تنها عضو یک گروه نبوده و ممکن است در چندین گروه عضویت داشته باشند، قرار گرفتن کاربران در گروه‌های مجزا، یک نقطه ضعف است [10]. از دیگر روش‌های پیشنهاد شده، خوشه‌بندی کاربران بر اساس ویژگی‌های آن‌هاست [11]. هرچند کاربران بر اساس علایق مشترک و شباهت‌هایشان در یک انجمن قرار می‌گیرند اما گروه‌های کشف شده بصورت جدا از هم بوده و چالش اساسی همچنان پابرجاست. در LBSN ها، اطلاعات کاملی از بازدیدهای کاربران شکل می‌گیرد که می‌تواند در خوشه‌بندی آن‌ها نقش به‌سزایی داشته باشد. یک چارچوب مناسب برای کشف انجمن‌های هم‌پوشا ارائه شده است که از شباهت کاربران و بازدیدهای آن‌ها استفاده کرده است [1]. نقص اساسی این چارچوب، انتخاب تصادفی و غیر هوشمند مراکز خوشه‌ها می‌باشد. همچنین، پژوهش-های مرتبط با این موضوع کار چندانی در ارائهٔ پروفایل انجمن‌ها و بازیابی خصوصیات شهرهای مختلف ارائه نکرده‌اند.

با عنایت به چالش مطرح شده در غیر هوشمند بودن انتخاب مراکز خوشه‌ها، در این مقاله یک رویکرد نوین جهت انتخاب مراکز خوشه‌ها توسط یکی از الگوریتم‌های فرامکاشفه‌ای ارائه می‌شود. یکی از معیارهایی که در خوشه‌بندی کاربرد دارد، استفاده از بازدیدهای انجام شده توسط کاربران است که یک منبع اطلاعاتی غنی برای رسیدن به شباهت حداکثری در خوشه‌ها خواهد بود. حاصل این تحقیق، کشف انجمن‌های هم‌پوشا از LBSN ها بصورت هوشمند و با دقت مناسب می‌باشد که به دنبال آن مشخصات شهرهای مختلف بازیابی می‌شوند.

ساختار کلی مقاله بدین صورت می‌باشد که در قسمت دوم، مروری مختصر از پیشینهٔ مرتبط با پژوهش مطرح می‌شود. در بخش سوم مسئله توضیح داده می‌شود. در بخش چهارم، جزئیات چارچوب پیشنهادی بیان می‌شود. در قسمت پنجم، نتایج حاصل از تحقیق ارائه می‌شود. در نهایت، نتیجه‌گیری کلی و مسیرهای باز تحقیق عنوان خواهد شد.

۲- پیشینهٔ پژوهش

کلیهٔ پژوهش‌های مرتبط با تحقیق در سه قسمت کلی قرار می‌گیرند که در این بخش، بطور مختصر به بررسی آن‌ها می‌پردازیم.

بخش اول- کشف انجمن مبتنی بر ساختار [12][13] [14] [15] : مهم‌ترین روش پیشنهاد شده «کشف انجمن‌های مبتنی بر مکان از گراف اجتماعی آنلاین با یال‌های حاشیه‌نویسی[9] شده توسط اطلاعاتی دربارهٔ محل‌های بازدید» می‌باشد [16].

بخش دوم- کشف انجمن مبتنی بر شباهت‌ها: تعریف دیگری که برای انجمن مطرح می‌شود، عبارتست از: «گروهی از کاربران که با هم شباهت دارند ولی با سایر اعضای شبکه شباهت ندارند.» بر اساس همین شباهت بین کاربران، انجمن‌ها با یک الگوریتم خوشه‌بندی کلاسیک کشف می-شوند. در سال ۲۰۰۹ دو رویکرد خوشه‌بندی مبتنی بر شباهت کاربران به منظور شناسایی انجمن‌ها در شبکهٔ Brightkite مطرح شد [11]. اما در ادامهٔ همین قسمت، پژوهش‌هایی با ارائهٔ معیارهای شباهتی جدید، توانستند به دقت بیشتری در نتایج دست یابند [17][18] [19] .

بخش سوم- کشف انجمن مبتنی بر روش ترکیبی: تمامی پژوهش‌های بررسی شده تاکنون بر کشف انجمن‌های جدا از هم تاکید داشته و کاربران را در گروه‌های مجزا خوشه‌بندی می‌کنند، از اینرو هیچ کدام در کشف انجمن‌های هم‌پوشا مناسب نیستند. به علاوه هیچ کدام از تحقیق‌های نام-برده شده، تلاشی به منظور ارائهٔ پروفایل و بازیابی خصوصیات شهرها انجام نداده‌اند. به منظور برطرف کردن این خلأ، یک چارچوب جدید در سال ۲۰۱۴ مطرح شد [3][1] . این چارچوب، یال‌های بین کاربران و محل‌های بازدید را بر اساس ویژگی آن‌ها خوشه‌بندی می‌کند. در این رویکرد، الگوریتم خوشه‌بندی مبتنی بر k-means کلاسیک و سلسله مراتبی ارائه شده است که تعیین مراکز خوشه‌ها بصورت تصادفی بوده و در حجم دادهٔ بالا ضعیف عمل می‌کند.

رویکرد پیشنهادی این مقاله در بخش سوم پژوهش‌های بررسی شده قرار دارد و به منظور ارتقاء آن می‌باشد. به همین منظور یک چارچوب هوشمند با استفاده از الگوریتم جستجوی هارمونی ارائه نمودیم که نقش این الگوریتم، انتخاب هوشمند مراکز خوشه‌ها با هدف افزایش شباهت درون خوشه‌ای می‌باشد.

---

[9] . Annotation

۳- تشریح مسئله

در این مقاله، انجمن به عنوان خوشه‌ای از یال‌ها (بازدیدهای کاربران) تعریف می‌شود که رئوس این یال‌های دوقسمتی شامل کاربر و محل بازدید می‌باشند. ما از بردار $U = (u_1,u_2,\dots,u_m)$ به عنوان مجموعهٔ کاربران و از بردار $V = (v_1,v_2,\dots,v_n)$ به عنوان مجموعهٔ دسته محل‌ها استفاده می‌کنیم. ماتریس بازدیدها بین کاربران و محل‌ها به نام M تعریف می‌شود. این ماتریس به بازدیدهایی که کاربر u_i از محل v_i انجام داده بستگی دارد. بدین ترتیب هر کاربر و هر دسته محل با یک بردار بازدید نمایان می‌شوند. کاربران و محل‌ها ویژگی‌های مستقل خود را دارا می‌باشند. براساس توضیحات بالا، مسئلهٔ کشف انجمن‌های هم‌پوشا در LBSN ها می‌تواند بصورت خوشه‌بندی یال‌های دوقسمتی باتوجه به ویژگی‌های کاربران و محل‌ها بیان شود. در تعریف انجمن، با یک مفهوم اساسی به نام «شباهت» مواجه می‌شویم. شکل کلی مسئله بصورت زیر فرموله شده است:

- ورودی‌ها:
  ۱. ماتریس بازدیدهای کاربران (M)
  ۲. ماتریس شباهتی کاربران
  ۳. ماتریس شباهتی محل‌های بازدید
  ۴. تعداد خوشه‌ها (k)
  ۵. پارامترهای الگوریتم فرامکاشفه‌ای
- خروجی: تعداد k انجمن هم‌پوشا

ویژگی‌های کاربران و محل‌های بازدید

ویژگی‌های درنظر گرفته شده در این پژوهش شامل دو گروه می‌باشند. گروه اول، ویژگی‌هایی که بر اساس ساختار شباهتی یال‌ها، یعنی بازدیدهای بین کاربران و دسته محل‌ها هستند (معیارهای شباهتی کاربر-محل بازدید و محل بازدید-کاربر). گروه دوم، شامل خصوصیات شباهتی کاربران و محل‌های بازدید با توجه به جنبه‌های اجتماعی می‌باشند (ویژگی شعاع گردش کاربر[10] و الگوی موقتی محل‌های بازدید). در مجموع ۴ شباهت در نظر گرفته شده به این شرح تعریف می‌شوند:

۱. شباهت کاربر-محل بازدید: در مجموعه دادهٔ مورد بررسی، برای شهر نیویورک 246 دسته محل بازدید و 913 کاربر شناسایی شده است. ما یک ماتریس 246*913 ساخته‌ایم که هر درایهٔ آن تعداد بازدیدهایی است که کاربر u_i از دسته محل v_j داشته است. بنابراین هر سطر این ماتریس نمایانگر یک بردار ویژگی کاربر-محل بازدید است. برای محاسبهٔ شباهت دو کاربر با بردارهای آن‌ها از شباهت کسینوسی استفاده می‌کنیم.

$$sim_{uv}(u_m,u_n)=\frac{\overrightarrow{u_m}\cdot\overrightarrow{u_n}}{\left\|\overrightarrow{u_m}\right\|\cdot\left\|\overrightarrow{u_n}\right\|} \qquad (1)$$

در رابطهٔ شمارهٔ 1، u_m ,u_n به بردارهای ویژگی دو کاربر اشاره دارد که شباهت بین دو کاربر با توجه به همین بردارها سنجیده می‌شود.

۲. شباهت محل بازدید-کاربر: با توجه به شباهت قبل، همان ماتریس ساخته می‌شود ولی هر سطر آن نشان دهندهٔ یک بردار ویژگی محل بازدید-کاربر است. همچنین برای محاسبهٔ شباهت بین دو دسته محل بازدید توسط بردارهای آن‌ها، از رابطهٔ شباهت کسینوسی استفاده می‌شود.

۳. شباهت شعاع گردش کاربر: یکی از معیارهای اساسی برای تشخیص سبک زندگی افراد مختلف، میزان جابجایی آن‌هاست. اینکه یک کاربر چقدر حرکت می‌کند با رابطهٔ شمارهٔ 2 معرفی می‌شود:

$$R_g=\sqrt{\frac{1}{n_i}[(x_i-x_m)^2+(y_i-y_m)^2]} \qquad (2)$$

رابطهٔ شمارهٔ ۲، برگرفته از فاصلهٔ اقلیدسی می‌باشد. همچنین x_i برابر طول جغرافیایی و y_i برابر عرض جغرافیایی محل بازدید شده توسط کاربر iام می‌باشد. دراین رابطه، مقدار x_m برابر میانگین طول جغرافیایی و y_m برابر میانگین عرض جغرافیایی محل‌هایی است که کاربر iام بازدید داشته است. همچنین n_i تعداد بازدیدهایی است که کاربر iام انجام داده است. برای محاسبهٔ شباهت بین کاربران با شعاع گردش، از رابطهٔ زیر استفاده می‌کنیم:

$$\mathrm{Sim}(u_m,u_n)=\frac{\min(R_g u_m,R_g u_n)}{\max(R_g u_m,R_g u_n)} \qquad (3)$$

در رابطهٔ شمارهٔ 3، نسبت کمترین به بیشترین مقدار شعاع گردش، شباهت بین دو کاربر u_m و u_n را تعیین می‌کند.

۴. شباهت الگوی موقتی محل بازدید[11]: باتوجه به اینکه مردم بازدیدهای مختلفی را در ساعات شبانه روز از محل‌های مختلف انجام می‌دهند، یک فرصت مقایسهٔ مناسب فراهم می‌آید. هر هفته به تعداد ساعات ۱۶۸(۲۴*۷) تفکیک می‌شود که هر عدد نشانگر یک ساعت از یک روز در هفته است. ماتریسی با تعداد ۱۶۸ ستون می‌سازیم و تعداد سطرهای آن به تعداد دسته محل‌ها بستگی دارد. هر سطر نشان می‌دهد که یک دسته محل در چه ساعاتی از هفته و شبانه روز بازدیدهای بیشتری داشته است.

بعد از تعیین ماتریس‌های شباهتی کاربران و محل‌ها، برای هرکدام دو ماتریس شباهتی خواهیم داشت. بنابراین جهت مقایسه و تصمیم‌گیری باید به یک ماتریس شباهتی برای کاربران و یک ماتریس شباهتی برای دسته محل‌ها دست یابیم. ترکیب شباهت‌ها مطابق روابط 4 و 5 خواهد بود:

$$\mathrm{sim}_u=\frac{1}{|f_u|}\sum \mathrm{sim}_u \qquad (4)$$

$$\mathrm{sim}_v=\frac{1}{|f_v|}\sum \mathrm{sim}_v \qquad (5)$$

در رابطهٔ شمارهٔ 4 و 5، $|f_u|$ و $|f_v|$ تعداد ویژگی‌های انتخاب شدهٔ کاربران و محل‌ها هستند. sim_u و sim_v شباهت‌های کاربران و محل‌ها می‌باشند.

٤- جزئیات پیاده‌سازی

در چارچوب پیشنهاد شده، ابتدا ویژگی‌های کاربران و محل‌های بازدید از مجموعهٔ داده انتخاب شده و شباهت بین کاربران و محل‌ها تعیین و سپس ادغام می‌شوند. در گام دوم، خوشه‌بندی یال‌ها با کمک الگوریتم جستجوی

[10] . Radius of Gyration (RG)

[11] . Venue Temporal Similarity

هارمونی انجام می‌گیرد. در نهایت یال‌های خوشه‌بندی شده با جایگزینی رئوس آن بازیابی می‌شوند. ایدهٔ اصلی این رویکرد، استفاده از الگوریتم جستجوی هارمونی جهت هوشمند شدن خوشه‌بندی یال‌ها می‌باشد.

در کشف انجمن‌ها، به دنبال این هستیم که شباهت درون خوشه‌ای حداکثر و شباهت با سایر انجمن‌ها حداقل باشد. از تابعی با هدف افزایش شباهت درون خوشه‌ای استفاده می‌کنیم:

$$\text{Object}=\max \sum_{j=1}^{k} \sum_{e_c \in C_j} \text{sim}(e_i, C_j) \qquad (6)$$

در رابطهٔ شمارهٔ 6، k تعداد انجمن‌ها، C مجموعهٔ انجمن‌های کشف شده، e_c یک یال از انجمن C_j و همچنین $\text{sim}(e_i,C_j)$ شباهت e_j و C_j می‌باشد.

با توجه به تابع هدف تعریف شده، نکتهٔ کلیدی اینست که شباهت بین یک یال و یک انجمن سنجیده شوند. شباهت بین یک جفت یال با شباهت بین رئوس مرتبط آن که شامل کاربر و محل بازدید است سنجیده می‌شود. در رابطهٔ شمارهٔ 7، تابع سنجش شباهت بین دو یال آمده است.

$$\text{sim}_{\text{edge}}(e_i, e_j) = F(\text{sim}_u(u_i, u_j), \text{sim}_v(v_i, v_j)) \qquad (7)$$

در رابطهٔ شمارهٔ 7، $\text{sim}_u(u_i,u_j)$ شباهت بین دو کاربر و $\text{sim}_v(v_i,v_j)$ شباهت بین دو محل بازدید است. تابع F در این مقاله بصورت میانگین هندسی تعریف شده و بصورت $\sqrt{\text{sim}_u * \text{sim}_v}$ می‌باشد. هر انجمن شامل مجموعه‌ای از یال‌هاست که بر مبنای رابطهٔ شمارهٔ 7 شباهت بین یال e_c و انجمن C_j بصورت زیر سنجیده می‌شود:

$$sim_{e_i C_j} = \frac{1}{|C_j|} \sum_{e_c \in C_j} sim_{edge}(e_i, e_j) \qquad (8)$$

در رابطهٔ شمارهٔ 8، $|C_j|$ تعداد یال‌های درون انجمن C_j می‌باشد. همانطور که در فرمول شمارهٔ 7 هم نشان داده شد، شباهت بین دو یال بر اساس شباهت رئوس آن تعریف می‌شود.

۴-۱- رویکرد پیشنهادی

بر اساس مطالبی که تاکنون از جزئیات پیاده سازی مسئله ذکر شد، ما باید الگوریتمی ارائه دهیم تا با توجه به معیارهای شباهتی تعریف شده، بتواند انجمن‌های هم‌پوشا را نمایان کند. الگوریتم‌هایی که تاکنون مطرح شدند بیشتر بر پایهٔ الگوریتم کلاسیک k-means بوده‌اند. یکی از عمده اشکالات اساسی این الگوریتم اینست که چون مراکز خوشه‌ها را بصورت تصادفی انتخاب می‌کند برای داده‌های حجیم عملکرد مناسبی ندارد. در این مقاله ما یک ابزار هوشمند به این الگوریتم اضافه نمودیم. بدین صورت که مبنای خوشه‌بندی همان الگوریتم k-means می‌باشد ولی نحوهٔ تعیین مراکز خوشه‌ها به کمک الگوریتم جستجوی هارمونی صورت می‌گیرد.

الگوریتم فرامکاشفه‌ای جستجوی هارمونی در سال ۲۰۰۱ توسط گیم و همکارانش ارائه شده است [20]. این الگوریتم، از فرایند جستجوی نوازنده‌ها به دنبال بهترین حالت از هارمونی برگرفته شده است. همانطور که آهنگ‌سازها بر اساس استانداردهای مربوط به زیبایی شناسی به دنبال هارمونی خوشایند می‌گردند، فرآیند بهینه‌سازی نیز براساس یک تابع هدف به دنبال بهترین جواب می‌گردد. گام مربوط به هرآلت موسیقی، خصوصیت زیبایی شناختی آن را تعیین می‌کند و مشابه مقدار تابع هدف (که بوسیلهٔ مجموعه‌ای از مقادیر که به هر متغیر تصمیم‌گیری اختصاص داده می‌شود) تعیین می‌گردد.

تابع برازندگی الگوریتم جستجوی هارمونی را مبتنی بر k-means تعریف می‌کنیم. با دقت در جدول شمارهٔ ۱، در خط ۱ با توجه به تعداد k، الگوریتم HS در فضای جستجو شروع به یافتن یال‌های بهینه کرده و آن‌ها را به عنوان مراکز خوشه‌ها به تابع برازش می‌دهد. در خطوط ۴ و ۵ شباهت بین کاربران و محل‌ها سنجیده می‌شود. سپس با استفاده از فرمول میانگین هندسی به یک مقدار واحد می‌رسد. تمامی شباهت‌های محاسبه شده در این تکرار درون یک بردار ذخیره می‌شود. با توجه به خط ۷، اگر مقدار این بردار شباهتی ذخیره شده بیشتر از تکرار قبلی بود، جایگزین آن می‌شود. هدف الگوریتم HS ماکزیمم کردن مجموع این بردار به منظور یافتن خوشه‌هایی با شباهت بیشتر است.

جدول ۱. شبه‌کد تابع برازندگی الگوریتم جستجوی هارمونی جهت کشف انجمن‌های هم‌پوشای مبتنی بر k-means

ورودی‌ها:

- E به عنوان لیستی از یال‌ها
- k، تعداد انجمن‌ها (خوشه‌ها)
- M_u، ماتریس شباهتی کاربران
- M_v، ماتریس شباهتی محل‌های بازدید

خروجی: C تعداد انجمن‌های هم‌پوشای کشف شده

۱. تعداد k یال را از مجموعه E توسط HS به عنوان مراکز خوشه‌ها انتخاب کن (C_j)
۲. برای هر مرکز خوشهٔ C_j انجام بده
۳. برای هر یال e_i در E انجام بده
۴. شباهت دو کاربر برای دو یال e_i و C_j را محاسبه کن
۵. شباهت دو محل بازدید برای دو یال e_i و C_j را محاسبه کن
۶. میانگین هندسی مقادیر به دست آمده از خطوط ۴ و ۵ را محاسبه کن و درون بردار شباهتی قرار بده
۷. اگر بردار شباهتی بدست آمده در این تکرار از تکرار قبلی بیشتر است آن را جایگزین کن
۸. مجموع بردار شباهتی را ماکزیمم کن

۴-۲- مجموعه داده‌ها

شبکهٔ اجتماعی مبتنی بر مکانی که ما آن را برای این مقاله هدف قرار دادیم، Foursquare می‌باشد [21]. این شبکه، دسترسی به بازدیدها و اطلاعات کاربران را محدود کرده است. بنابراین، باتوجه به اینکه حدود۲۰ تا ۲۵ درصد کاربران Foursquare، بازدیدهای خود را از طریق توئیتر به اشتراک می‌گذارند، برای بدست آوردن داده‌های مورد نیاز، از واسطهای شبکهٔ توئیتر استفاده شده است. مجموعه دادهٔ در دسترس ما، در طول ۳ ماه و از دو شهر نیویورک و توکیو جمع‌آوری شده است. این مجموعه داده برای شهر نیویورک شامل ۹۱۳ کاربر، ۲۴۶ دسته محل، و ۵۰۷۶۵ بازدید می‌باشد. همچنین داده‌های شهر توکیو شامل ۲۰۴۱ کاربر، ۲۲۲ دسته محل و ۶۶۳۱۵ بازدید می‌باشد.

۵- نتایج پیاده‌سازی

هدف از پروفایل انجمن‌ها اینست که ببینیم انجمن کشف شده به چه چیزی شباهت دارد. پروفایل هر انجمن با فراداده‌های اعضای آن انجمن یعنی کاربران و محل‌های بازدید تعیین می‌شود. فراداده برای کاربران می‌تواند شعاع گردشی و برای محل بازدیدها می‌تواند دستهٔ آن‌ها باشد. برای مشخص شدن یک انجمن در ابتدا اهمیت هر کاربر و دسته محل را با توجه به درجهٔ دخالتشان محاسبه می‌کنیم. سپس انجمن را با ساخت یه بردار ویژگی که خصوصیات آن انجمن را بطور خلاصه دارد، توصیف می‌کنیم.

بطور خاص، اهمیت کاربر U_m در انجمن C_j با درصد تعداد بازدیدهای آن کاربر در انجمن مرود نظر مشخص شده و اهمیت دستهٔ محل بازدید در یک انجمن با درصد تعداد بازدیدهای صورت گرفته از آن دسته محل معرفی می‌شود. یک موجودیت درصورتی معنادار محسوب می‌شود که اهمیت آن از یک حد آستانه بیشتر باشد. در این مقاله حد آستانه را برای کاربران ۰,۰۰۱ و برای دسته محل‌ها ۰,۰۲ درنظر گرفتیم. دلیل درنظر گرفتن این حدود آستانه، نمایش تعدادی قابل توجه از کاربران و دسته-محل‌های بازدید است و این مقادیر بصورت تجربی تعیین شدند. لیستی از کاربران و دسته محل‌های مهم بصورت زیر نمایش داده می‌شود:

$p_{C_j}=\{<f_{u_1}, e_{u_1}>,\ldots,<f_{u_m}, e_{u_m}>\ldots<f_{v_1}, e_{v_1}>,\ldots,<f_{v_n}, e_{v_n}>\}$

در این تعریف، هر سطر $<f_{u_m}, e_{u_m}>$ و $<f_{v_n}, e_{v_n}>$ مقادیر مربوط به ویژگی حالت کاربر یا محل بازدید را نشان می‌دهد.

یکی از انجمن‌های شناسایی شده در شهر نیویورک را برای شبکهٔ Foursquare در نظر گرفتیم (ورودی تعداد خوشه‌ها را ۳۰ فرض کردیم). این انجمن شامل ۳۸۸ کاربر و ۷۰ دسته محل بازدید و همچنین تعداد ۴۹۸۵ بازدید می‌باشد. بر مبنای تعریف بالا مبنی بر اهمیت اعضای یک انجمن، تعداد ۱۸۰ کاربر و تعداد ۶ دسته محل بازدید در این انجمن مهم هستند. این انجمن به صورت زیر پروفایل می‌شود:

پروفایل انجمن مورد نظر = {<خانه،۰,۶۴۷۷> <ساختمان‌های دولتی،۰,۰۲۲۷> <شام،۰,۰۲۱۳> <اغذیه،۰,۰۲۱۹> <کافه،۰,۰۲۱۱> < فروشگاه کتاب ،۰,۰۲۵۵> <شعاع گردشی کاربران مهم،۰,۰۶۱۲>}

با توجه به پروفایل این انجمن، دسته محل خانه دارای بیشترین اهمیت می‌باشد، پس بنابراین انجمن کشف شده به همین گروه تعلق دارد.

مشخصات شهرها

بعد از شناسایی انجمن‌ها و پروفایل آن‌ها به سراغ بازیابی ویژگی و جنبه-های شهرهای مختلف می‌رویم. در این مقاله دو شهر نیویورک و توکیو مورد تحلیل قرار گرفته و انجمن‌های آن بازیابی شده‌اند. تعداد خوشه‌ها را ۱۰ فرض کردیم و این تعداد خوشه را در ۸ گروه کلی قرار دادیم تا یک بررسی و مقایسهٔ منطقی داشته باشیم. سپس پروفایل شهرهای مختلف را نمایش داده و مشخصات دو شهر نیویورک و توکیو را مقایسه می‌کنیم. نتایج به دست آمده در جداول ۲ و ۳ نشان داده شده و نام‌گذاری هر طبقه بر اساس مهم‌ترین دسته محل‌های آن گروه صورت می‌گیرد. اعداد هر سطر شامل تعداد بازدیدهای انجام شده از گروه، تعداد کاربران گروه، نسبت تعداد کاربران گروه به تعداد کل کاربران Foursquare در شهر مربوطه و تعداد دسته محل‌های بازدید آن گروه می‌باشد.

جدول ۲. مشخصات استخراج شده برای شهر توکیو

توکیو				
دسته‌محل‌ها	نسبت کاربران به کل کاربران	کاربران	بازدیدها	
۱۸۰	٪ ۶۲	۱۲۸۱	۸۵۸۶	خارج از خانه
۱۴۶	٪ ۱۸	۳۷۶	۲۷۰۵	حمل و نقل عمومی
۱۵۶	٪ ۶۸	۱۳۹۱	۳۴۲۵۳	نوشیدنی/ کافه
۸۸	٪ ۶۳	۱۳۰۰	۷۵۱۴	مراکز عمومی
۸۵	٪ ۲۴	۴۹۰	۱۸۵۷	آموزش
۹۸	٪ ۶۹	۱۴۱۳	۶۳۳۷	غذا
۹۶	٪ ۴۵	۹۳۳	۳۹۷۳	خانه
۵۹	٪ ۴۷	۹۶۷	۱۰۸۵	تفریح

جدول ۳. مشخصات استخراج شده برای شهر نیویورک

نیویورک				
دسته‌محل‌ها	نسبت کاربران به کل کاربران	کاربران	بازدیدها	
۱۶۷	٪ ۹۴	۸۶۲	۱۰۷۵۱	خارج از خانه
۸۵	٪ ۳۰	۲۸۳	۲۱۰۸	حمل و نقل عمومی
۱۳۷	٪ ۸۲	۷۵۲	۱۱۸۵۰	نوشیدنی/ بار
۸۵	٪ ۴۳	۳۹۵	۳۱۹۸	مراکز عمومی
۹	٪ ۱۰	۱۰۰	۱۹۴	آموزش
۱۷۰	٪ ۷۳	۶۷۳	۱۱۱۸۱	غذا
۱۲۹	٪ ۴۱	۳۸۰	۶۲۶۸	خانه
۱	٪ ۷۲	۶۶۴	۵۲۱۴	تفریح

با مقایسهٔ نتایج به دست آمده، مشخصات و جنبه‌های مختلف دو شهر توکیو و نیویورک نمایان می‌شود. برای مثال بالا بودن درصد حضور در بیرون از خانه کاملا طبیعی و قابل توجیح است، زیرا اغلب مردم زمان زیادی را بیرون از خانه به عنوان کار، رانندگی، و غیره سپری می‌کنند. اما مقایسه بین سایر گروه‌ها نتایج جالب وشگفت‌انگیزی را به همراه دارد. برای نمونه در گروه نوشیدنی (بار، کافه و غیره) کاربران شهر نیویورک نسبت به کاربران شهر توکیو حضور بیشتری داشته و شاید بتوان این موضوع را با فرهنگ کشور آمریکا مرتبط دانست. اما با مقایسهٔ گروه غذا برای این دو شهر مشاهده می‌کنیم که اعضای بیشتری از هر دو شهر به این گروه علاقه دارند و اختلاف چندانی بین دو شهر مشاهده نمی‌شود. کاملا منطقی است که نیاز طبیعی به تغذیه برای تمامی افراد بوده و تفاوت چندانی بین فرهنگ و جغرافیا وجود ندارد. اما در تمام دنیا، مردم کشور ژاپن به سخت-کوشی و تلاش شهرت دارند، با بررسی گروه تفریح (پارک، کلوپ و غیره)

این مورد قابل فهم خواهد بود. کاربران شهر نیویورک ۷۲ درصد و کاربران شهر توکیو ۴۷ درصد اهل تفریح هستند. شاید بتوان این موضوع را اهمیت ندادن مردم توکیو به تفریح و بیشتر مشغول کار بودن، دانست.

بر مبنای جنبه‌ها و ویژگی‌های شهرها، نهادهای مختلف می‌توانند بر روی سرویس‌دهی و توسعهٔ کار خود تمرکز کرده و حداکثر سود را کسب نمایند. برای مثال اگر یک رستوران زنجیره‌ای قصد احداث یک شعبهٔ جدید را داشته باشد، می‌تواند ابتدا علایق و مشخصات یک شهر را بررسی کرده و با توجه به آن، اقدام به احداث نماید. با این پیش‌بینی و امکان‌سنجی، درصد شکست سرمایه‌گزاری پایین آمده و احتمال موفقیت بالاتر می‌رود.

۶- نتیجه‌گیری و مسیرهای باز تحقیق

در این مقاله، با تکیه بر اطلاعات به دست آمده از بازدیدهای کاربران و همچنین ویژگی محل‌ها و کاربران، یک چارچوب هوشمند جهت تشخیص انجمن‌های هم‌پوشا در LBSN ها ارائه شده است. این رویکرد جدید با استفاده از الگوریتم‌های فرامکاشفه‌ای مانند جستجوی هارمونی، می‌تواند دقت بالاتری در ارائهٔ انجمن‌ها داشته باشد. در مقایسهٔ این رویکرد با رویکردهای پیشین به این نکته رسیدیم که نقطهٔ عطف پژوهش ما تعیین هوشمند مراکز خوشه‌ها می‌باشد. کشف انجمن‌های هم‌پوشا از LBSN ها و بازیابی مشخصات شهرهای مختلف، در بسیاری از امکان‌سنجی‌های تجاری، بازاریابی و تبلیغات گروهی کاربرد خواهد داشت.

از مسیرهای باز این تحقیق می‌توان به طراحی یک چارچوب چند هدفه به منظور کشف انجمن‌های هم‌پوشا اشاره نمود که مراکز خوشه‌ها و تعداد انجمن‌ها بصورت هوشمند و با افزایش شباهت درون خوشه‌ای صورت می‌گیرد.

منابع

[1] Z. Wang, D. Zhang, X. Zhou, D. Yang, Z. Yu and Z. Yu, "Discovering and profiling overlapping communities in location-based social networks," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 44, no. 4, pp. 499-509, 2014.

[2] Y. Zheng, "Location-based social networks: Users," in *In Computing with spatial trajectories*, New York, Springer, 2011, pp. 243-276.

[3] Z. Wang, X. Zhou, D. Zhang, B. Guo and Z. Yu, "Community Detection and Profiling in Location-Based Social Networks," in *Creating Personal, Social, and Urban Awareness through Pervasive Computing*, Hershey, IGA Global, 2013, pp. 158-175.

[4] Y. Zheng and X. Xie, "Location-based social networks: Locations.," in *In Computing with spatial trajectories* , New York, Springer , 2011, pp. 277-308.

[5] J. Bao, Y. Zheng, D. Wilkie and M. Mokbel, "Recommendations in location-based social networks: a survey," *GeoInformatica*, vol. 19, no. 3, pp. 525-565, 2015.

[6] J. Scott, Social network analysis, Los Angeles, London, New Delhi, Singapore, Washington: Sage, 2012.

[7] C. Tantipathananandh and T. Berger-Wolf, "Finding communities in dynamic social networks," in *In Data Mining (ICDM), 2011 IEEE 11th International Conference on*, Vancouver,BC, 2011.

[8] S. Fortunato, "Community detection in graphs," *Physics reports* , vol. 486, no. 3-5, pp. 75-174, 2010.

[9] Y. Ahn, J. Bagrow and S. Lehmann, "Link communities reveal multiscale complexity in networks," *Nature*, vol. 466, no. 7307, p. 761–764, 2010.

[10] J. Cruz, C. Bothorel and F. Poulet, "Entropy based community detection in augmented social networks," in *Computational Aspects of Social Networks (CASoN), 2011 International Conference on*, Salamanca, 2011.

[11] Y. C. H. Y. J. Zhou, "Graph clustering based on structural/attribute similarities," *Proceedings of the VLDB Endowment*, vol. 2, no. 1, pp. 718-729, 2009.

[12] Z. Cheng, J. Caverlee, K. Lee and D. Sui, "Exploring Millions of Footprints in Location Sharing Services," in *ICWSM*, 2011.

[13] M. Vasconcelos, S. A. J. Ricci, F. Benevenuto and V. Almeida, "Tips, dones and todos: uncovering user profiles in foursquare," in *The fifth ACM international conference on Web search and data mining*, 2012.

[14] N. Li and G. Chen, "Analysis of a location-based social network.," in *In Computational Science and Engineering*, 2009.

[15] K. Wakita and T. Tsurumi, "Finding community structure in mega-scale social networks," in *The 16th international conference on World Wide Web*, 2007.

[16] V. Blondel, J. Guillaume, R. Lambiotte and E. Lefebvre, "Fast unfolding of communities in large networks.," *Journal of statistical mechanics: theory and experiment*, vol. 2008, no. 10, pp. 10008-10020, 2008.

[17] A. Noulas, S. Scellato, C. Mascolo and M. Pontil, "An Empirical Study of Geographic User Activity Patterns in Foursquare," in *ICwSM*, 2011.

[18] S. Scellato, C. Mascolo, M. Musolesi and V. Latora, "Distance Matters: Geo-social Metrics for Online Social Networks," in *WOSN*, 2010.

[19] S. Scellato, A. Noulas, R. Lambiotte and C. Mascolo, "Socio-Spatial Properties of Online Location-Based Social Networks," in *ICWSM*, 2011.

[20] Z. Geem, J. Kim and G. Loganathan, "A new heuristic optimization algorithm: harmony search," *Simulation*, vol. 76, no. 2, pp. 60-68, 2001.

[21] D. Yang, D. Zhang, V. W. Zheng and Z. Yu, "Modeling User Activity Preference by Leveraging User Spatial Temporal Characteristics in LBSNs," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 45, no. 1, pp. 129-142, 2015.