کاوش الگوهای خط‌سیر در اشیا متحرک در شبکه‌های جاده‌ای

نیلوفر میرزائی چهارده[۱] بهروز مینایی بیدگلی[۲] محمدرضا عباسی‌فرد[۳]

۱- فارغ التحصیل- دانشکده مهندسی کامپیوتر- دانشگاه غیرانتفاعی رجا- قزوین- ایران
Nil_Mirzaiee@yahoo.com
۲- دانشیار- دانشکده مهندسی کامپیوتر- دانشگاه علم و صنعت - تهران- ایران
B_Minaei@iust.ac.ir
۳- استادیار- دانشکده مهندسی کامپیوتر- موسسه غیرانتفاعی ادیبان - گرمسار- ایران
Abbasifard@adiban.ac.ir

چکیده: به کارگیری الگوکاوی بر روی داده‌های مرتبط با مسیرها و جاده‌ها، دستاوردهای باارزش و اثرگذاری را در حوزه‌های مختلفی، همچون راه و شهرسازی، حمل و نقل و حتی پیش‌بینی سرویس‌های اجتماعی به ارمغان آورده است. در این مقاله به کاوش الگوهای خط‌سیر در اشیا متحرک در شبکه‌های جاده‌ای پرداخته شده است. تا بتوان به کمک آن الگوهای مناسبی را استخراج نماییم. ازجمله چالش‌های موجود در تکنیک‌های خوشه‌بندی در بحث الگوکاوی، پیدا کردن الگوهای متوالی همراه با زمان اجرای پایین می‌باشد. برای حل این مشکل، از تکنیک‌های خوشه‌بندی برای معرفی یک الگوریتم پیشنهادی به نام BFEs-Enhanced استفاده شده است، که با استفاده از آن می‌توان الگوهای Flock متوالی و معنی‌داری را که دارای بازه زمانی پایینی می‌باشند، استخراج نمود. نتایج شبیه‌سازی نشان می‌دهد که، روش پیشنهادی به دلیل استفاده از معیارهای مناسب برای خوشه‌بندی که شامل: حداکثر فاصله ثابت و حداقل تعداد می‌باشد و اضافه کردن معیار جدیدی به نام حداقل دوره زمانی به الگوریتم پیشنهادی و از طرف دیگر، در روش پیشنهادی علاوه بر معیار فاصله که الگوریتمBFE-Enhanced داشته است، معیار جهت نیز برای بهبود دقت عملکرد به الگوریتم پیشنهادی اضافه شده است. که در نتیجه، الگوریتم پیشنهادی، بازه زمانی را کاهش و تعداد الگوهای معنی‌دار و متوالی را افزایش می‌دهد.

واژه‌های کلیدی: الگوکاوی خط‌سیر، اشیا متحرک، شبکه‌های جاده‌ای، الگوریتم‌های خوشه‌بندی، الگوریتم BFE-Enhanced

نوع مقاله: پژوهشی

DOI: 10.52547/jiaeee.20.3.125

تاریخ ارسال مقاله: ۱۴۰۰/۱۰/۱۶
تاریخ پذیرش مشروط مقاله: ۱۴۰۱/۰۴/۱۳
تاریخ پذیرش مقاله: ۱۴۰۱/۰۷/۱۳
نام نویسنده‌ی مسئول: دکتر بهروز مینایی بیدگلی
نشانی نویسنده‌ی مسئول: ایران – تهران – نارمک – دانشگاه علم و صنعت ایران – دانشکده مهندسی کامپیوتر

۱- مقدمه

اخیرا، با پیشرفت تکنولوژی‌هایی همچون: سیستم‌های مبتنی بر مکان، از وجود هزاران خط‌سیر[۱] گوناگونی که توسط اشیا متحرک[۲] مانند: افراد، وسایل نقلیه، حیوانات و پدیده‌های طبیعی در حوزه‌های مختلف شهری و به صورت آگاهانه و یا غیرآگاهانه تولید می‌شوند، استفاده می‌نمایند. خط‌سیرهای مکانی–زمانی[۳]، اطلاعات ارزشمندی در ارتباط با اشیا متحرک و موقعیت آنها به محققان و پژوهشگران ارائه می‌کنند. که آنها، می‌توانند برای پژوهش‌های منظم و توسعه تکنولوژی جدید محاسباتی برای کارهایی همچون: پردازش، بازیابی و کاوش داده‌های خط‌سیر در اشیا متحرک مورد استفاده قرار گیرند.

به طور کلی محاسبات خط‌سیرهای مکانی-زمانی، اهمیت ویژه‌ای را در انواع پژوهش‌ها داشته است و در زمینه‌های مختلف، مانند: برنامه‌ریزی و مدیریت شهری، زیست‌شناسی، جامعه‌شناسی، جغرافیا، هواشناسی و علوم کامپیوتر نظیر: پایگاه‌داده، GIS، داده‌کاوی[۴] نیز توجه دانشمندان را به خود جلب نموده است، و نمی‌توان نقش این نوع از فناوری‌های جدید را در پیشرفت حوزه‌های مختلف نادیده گرفت. این حوزه‌ها همچون: شبکه‌های جاده‌ای می‌باشند، که در آنجا می‌توان به اهمیت انجام کاوش الگوهای خط‌سیر[۵] پی‌برد. بنابراین کشف الگوهای خط‌سیر یک نقش مهمی را در اشیا متحرک بازی می‌کنند، و استخراج و تحلیل داده‌های خط‌سیر می‌تواند خدمات بیشتری را برای کاربران فراهم نماید. تا به الان، کاوش داده‌های خط‌سیر اساسا بر روی دو جهت خوشه‌بندی خط‌سیر[۶] و کاوش الگوهای خط‌سیر متمرکز شده است.[۶]

از جمله چالش‌های موجود در تکنیک‌های خوشه‌بندی، معیار خوشه‌بندی می‌باشد، به این صورت که اکثرا خوشه‌بندی تنها براساس معیار فاصله انجام می‌شود. به این معنا که چند مسیر با توجه به معیار فاصله، به یکدیگر نزدیک هستند، بنابراین درون یک خوشه قرار می‌گیرند، اما مسیرهای موجود، درون هر خوشه نسبت به مسیرهای خوشه دیگر، از نظر معیار فاصله به هیچ عنوان شباهتی ندارند. همچنین برای غلبه بر این چالش، الگوریتم پیشنهادی باید از طریق قراردادن معیار دیگری در کنار معیار فاصله، مسیرها را خوشه‌بندی و الگوریتم‌های متوالی و معنی‌دارای را از آن استخراج نماید.

با توجه به چالش و مطالب مطرح شده در این قسمت، در این پژوهش سعی شده است، معیار دیگری به نام جهت را نیز به تکنیک‌های خوشه‌بندی اضافه نماییم، تا خوشه‌بندی با دقت بالاتری انجام شود. همچنین سعی شده است که در الگوریتم پیشنهادی الگوهای Flock معنی‌دار و به صورت متوالی از داده‌های خط‌سیر موجود به دست آورده شود. مطالبی که در ادامه این مقاله خواهیم داشت، ابتدا پیشینه پژوهش، سپس روش پیشنهادی که به تفصیل بیان شده است و پس از آن به بیان یافته‌ها و در نهایت به نتیجه‌گیری مطالب ارائه شده در مقاله و بیان کارهای آینده خواهیم پرداخت.

۲- پیشینه پژوهش

کاوش الگوهای خط‌سیر با داده‌های GPS، به طور گسترده برای پیدا-کردن الگوهای حرکتی اشیا متحرک مورد بررسی قرار گرفته است. که ابتدا در [۲۰,۵] درباره تکنیک‌های کاوش خط‌سیر سنتی بحث شده است، که تنها از موقعیت‌یاب GPS استفاده می‌کنند و سپس با مطالعه بر روی سایر الگوریتم‌ها بررسی بیشتری انجام شده است. [۱۴-۱۸-۲۱-۲۲] همچنین، با استفاده از مکان‌یابی معنایی[۷]، الگوهای خط‌سیر را نیز پیدا می‌کنند.

الگوریتم‌های کاوش خط‌سیر سنتی، معمولا خوشه‌هایی از تراکنش‌های مشابه را برای پیدا کردن خط‌سیرهای عمومی و پرکاربرد استفاده می‌کنند. برای مثال: یک الگوریتم خوشه‌بندی در [۵] پیشنهاد شده است. که نه تنها گروهی از خط‌سیرهای مشابه را، همچنین خوشه‌هایی با دنباله‌های مشابه، را نیز کشف می‌کنند. مسئله پیداکردن الگوهای خط‌سیر مختلف و جالب، اخیرا توسط [۱۴] بررسی شده است، که یک الگوریتم خوشه‌بندی، برای پیدا کردن الگوهای خط‌سیر ارائه نموده است. در [۱۲] به بررسی مسیرهای وسیله نقلیه با توجه به فواصل بین بردارهای ویژه بخش خط‌سیر با استفاده از روش خوشه‌بندی سلسله‌مراتبی پرداخته شده است. در[۵] خوشه‌بندی با استفاده از الگوریتم Two-Phase انجام شده است. که ابتدا، پارتیشن-های خط‌سیر را با استفاده از اصل MDL و سپس خوشه‌ها و بخش-هایی از خط‌سیر را، با استفاده از الگوریتم‌های خوشه‌بندی بخش‌های خطی، انجام می‌دهند. در [۲۴] دیدگاه اصلاح فیلترها، برای کشف-کردن خط‌سیرها، در پایگاه داده‌های خط‌سیر مورد بررسی قرار گرفته است.

در [۸]، یک پارتیشن‌بندی نهایی برای تشخیص خط‌سیرهای خارجی تعیین شده است. مخصوصا یک دیدگاهی، که ترکیبی براساس مسافت و براساس چگالی، برای توسعه و شناسایی ناهنجاری-ها می‌باشد، استفاده شده است. در [۲۵] سه گروه از مجموعه الگوها در پایگاه داده‌ای متحرک که شامل: خوشه‌های حرکتی[۸]، پرس‌و-جوهای Convoy و الگوهای Flock می‌باشند، را نام برده است. هم خوشه‌های حرکتی[۴] و هم پرس‌وجوهای Convoy [۲۶] در این مسئله، که آنها هر دو براساس الگوریتم خوشه‌بندی هستند، مشترک می‌باشند. الگوهای Flock در واقع مجموعه‌ای از اشیایی که با یکدیگر برای یک مدت زمان مشخص حرکت می‌کنند تعریف می‌شوند[۱۳]. همچنین الگوریتم‌هایی نیز، براساس چگالی وجود دارند، از قبیل: DBSCAN [۱۹]، این الگوریتم از پارامترهایی مثل: Minpts و بیشترین مسافت[۹] (ε) برای محتویات یک خوشه استفاده می‌کند. از معایب DBSCAN می‌توان به مدیریت کردن پایگاه داده‌های بزرگ و کیس‌هایی که پیچیدگی آنها بدتر از DBSCAN با $O(n^2)$ است، اشاره نمود[۱۷]. به تازگی در [۲] به گسترش الگوریتم‌های استخراج الگوهای Flock خط‌سیرگرا کلاسیک، برای کشف الگوهای Flock

متحرک برای ترانزیت مسافران، حسابداری برای محدودیت شبکه‌های حمل و نقل چند وجهی را مورد بررسی قرار داده است.

آنها استفاده از دیسک‌ها با یک شعاع مشخص برای تشخیص گروهی از خط‌سیرهای حرکتی با یکدیگر در همان جهت را معرفی می‌کنند. تمام خط‌سیرهایی که هر کدام در کنار دیسک مورد نظر در یک فاصله زمانی خاص قرار می‌گیرند، به عنوان یک الگو نامزد[۱۰] بررسی می‌شوند. در [۱] کشف Flockهای ثابت نشان داده شده است. خود این مساله، یک مشکل Np-Hard است. اخیرا در [۲۵] یک الگوریتم On-Line برای پیدا کردن الگوهای حرکتی Flock معرفی شده است، که اصطلاحا آن را BFE می‌نامند، این الگوریتم به عنوان راه‌حلی دقیق، برای حل مسائلی همچون الگوهای Flock در زمان‌های چندجمله‌ای معرفی شده است. اخیرا در [۱۱] یک الگو حرکتی Flock تعریف شده است، و آن را به صورت الگوریتمی تطابق‌پذیر[۱۱] بیان نموده است، که براساس آن نظریه، دارای وابستگی مکانی و زمانی می‌باشد.

برای غلبه بر مشکل الگوهای تکراری، استخراج الگوهای Maximal و Closed پیشنهاد شده است. در [۱۶,۱۰] تکنیک‌های فوق‌الذکر برای حوزه‌های مختلفی همچون بیوانفورماتیک، GIS و بازاریابی و مخصوصا برای پیاده‌سازی استخراج مجموعه اقلام تکراری[۱۲] FIMI [۲۳,۶] به صورت موفقیت‌آمیزی اجرا شده است. اولین بار، پیدا کردن الگوهای Flock معرفی شده است در [۱۹,۷]، به هر حال، آنها نظریه‌ای درباره دوره‌های زمانی را بررسی نکرده‌اند.

برای تشخیص شباهت زیاد Flockها، تنها دو متغیر استفاده شده است. اولین متغیر ماکزیمم فاصله ثابت[۱۳] در طول حرکت اشیا ($\in$) است. که به عنوان شعاع دیسک آن را در نظر گرفته است. دومین متغیر، حداقل تعداد[۱۴] اشیا متحرک(μ) است، که هرکدام از این دو متغیر باید در کنار آن دیسک قرار بگیرند. بعد از مدتی، در [۴] یک متغیر جدیدی، به نام حداقل دوره زمانی[۱۵] (δ) معرفی کردند. که به عنوان یک پارامتر جدید در Flock مورد بررسی قرار گیرد. اخیرا یک الگوریتمی به نام SPARROW که از ترکیب مفاهیم DBSCAN و تکنیک‌های هوشمندانه Swarm در [۹] به وجود آمده است. این الگوریتم، برای استفاده در داده‌های Big data مناسب است. به تازگی در [۲۷]یک چهارچوب کلان داده را برای پیش‌بینی داده‌های مسیر جریان با بهره‌برداری از الگوهای استخراج شده از مسیرها را ارائه می‌کند که پیش‌بینی بلندمدت دقیق با تاخیر کم را بررسی می‌کند.

در [۱] ساختارهای مختلفی برای استخراج الگوهای تکراری تعریف شده‌اند. یکی از معمولی‌ترین ساختارها براساس Support آنها است، که در هرکدام از مجموعه داده‌هایی با تکرار بالا، با یک Threshold معین تعیین شده است. همچنین به تازگی در [۳] به توسعه روشی که مسیر را براساس تغیر حالت حمل و نقل در زمان واقعی بدون قوانین پیش‌فرض تقسیم کند می‌پردازد و همچنین از ماتریس‌های حالت‌گذار (TSM) برای تشخیص خودکار نقطه تغیر حالت انتقال در مسیر استفاده می‌کند.

۳- روش پیشنهادی

از مهمترین چالش‌ها در تکنیک‌های خوشه‌بندی در مبحث الگوکاوی، پیداکردن الگوهای متوالی با زمان اجرای پایین می‌باشد. به همین دلیل روش پیشنهادی سعی در ارائه یک الگوریتم مناسب که با استفاده از تکنیک‌های خوشه‌بندی به استخراج الگوهای Flock متوالی و معنی‌دار که در بازه زمانی پایین به دست بیایند، نموده است. بنابراین روش پیشنهادی باید مشکلاتی از جمله زمان اجرای بالا، الگوهای بی‌معنی و مقطعی، و نیاز به استفاده از سایر الگوریتم‌ها برای مرحله هرس کردن خوشه ها را نیز برآورده سازد.

روش پیشنهادی موجود در این مقاله را اصطلاحا BFEs-Enhanced می‌نامند، که این روش برگرفته از الگوریتم BFE-Enhanced می‌باشد و به دلیل پیدا کردن الگوهای متوالی و پی در پی[۱۶] این نامگذاری انجام شده است. پیداکردن الگوهای حرکتی Flock در الگوریتم پیشنهادی همچون: الگوریتم BFE-Enhanced و Sparrow در [۱۷] می باشد، که در ۴ مرحله به صورت زیر انجام می‌شود:

۳-۱- به دست‌آوردن یک مجموعه نهایی از خوشه-های معتبر در هر بازه زمانی

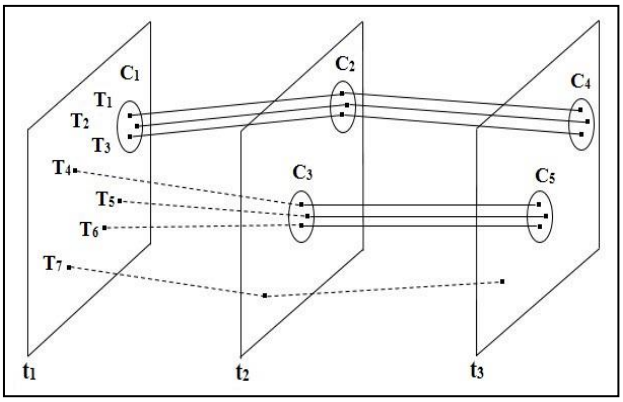
مجموعه داده‌های مورد استفاده در این پیاده‌سازی، تنها بخشی از مجموعه داده‌های عظیم پایگاه‌داده GeoLife می‌باشد. در این پژوهش، ابتدا به صورت آزمایشی، فواصل زمانی را به مدت، پنج‌ماه مورد آزمایش قرار داده شد، به این صورت که، داده‌ها از تاریخ آگوست ۲۰۰۷ تا دسامبر ۲۰۰۷ در نظر گرفته شده‌اند؛ و جدولی برای هر بازه زمانی ایجاد شد. سپس برای آزمون نهایی، فواصل زمانی را برای دوازده‌ماه مورد آزمون قرار داده شد، به این صورت که از تاریخ آوریل ۲۰۰۷ تا مارچ ۲۰۰۸ مورد بررسی قرار گرفتند.

۳-۲- ایجاد یک نسخه از تراکنش‌ها در مجموعه داده‌های خط‌سیر براساس دیسک‌های ملاقات شده به وسیله هر خط‌سیر

برای درک کلی مجموعه داده‌های مکانی-زمانی این گونه می‌توان بیان نمود که، مجموعه داده‌های مکانی-زمانی شامل: اطلاعاتی برای مکان یک موجودیت در یک زمان خاص هستند. هر ورودی یک مجموعه داده شامل: ویژگی‌های یک نقطه از یک خط‌سیر را منعکس می‌کند. که برای تجزیه و تحلیل روش‌های مختلف در مجموعه داده-ها توانمند می‌باشند. در این مقاله، فرض شده است که مجموعه داده مکانی- زمانی، شامل دو فیلد است:

- یک ID خط‌سیر، که متعلق به یک نقطه است، زمانی که مختصات X,Y یک مکان را اندازه‌گیری می‌کند. اینکه بتوانند از الگوریتم استخراج الگوهای تکراری استفاده نمایند، باید ID خط‌سیر بتواند برای شناسایی تراکنش‌های برابر و تعیین ID اقلامی که هرکدام، مجموعه‌ای از اطلاعاتی شامل: مکان و زمان برای هر نقطه است، را استفاده نماید.
- یک ID منحصر به فرد برای هر دیسک تولید شده در اولین گام نیز ضمیمه شده است. علاوه بر این، اطلاعاتی درباره هر کدام از خط‌سیرهای ملاقات شده از یک دیسک در فواصل زمانی مخصوص نیز در یک جدول به صورت جدا ذخیره شده است. بنابراین، برای بدست آوردن یک نسخه‌ای از تراکنش‌های خط‌سیر، باید زمان و مکان برای یک نقطه از دیسک با ID آن دیسک برابر شود.

یک دیسک به عنوان یک محدوده مخصوص در مکان و زمان هر خط‌سیر می‌تواند، برطبق دیسک‌هایی که هر کدام از آنها در طول زمان زندگی[17] آنها ملاقات شده‌اند، ترجمه شود. این مفهوم به صورت مثالی، در شکل شماره ۱ به صورت زیر نشان داده شده است.



شکل (۱): مثالی از یک خط‌سیر در داخل مجموعه‌داده[۱]

با توجه به شکل شماره۱، یک مجموعه‌ای شامل: هفت خط-سیر(T_i) را نشان داده است، بنابراین پنج دیسک می‌تواند از طریق Lifetime مجموعه داده (C_i) تشخیص داده شود. که در جدول شماره ۱ هرکدام از دیسک‌هایی که برای هر خط‌سیر در بازه زمانی مشخصی (t_j) ملاقات انجام شده است، را نیز در این شکل نشان داده شده است.

اگر جدول شماره۱، را به عنوان یک پایگاه داده تراکنشی در نظر بگیریم، می‌توانیم از الگوریتم استخراج الگوهای تکراری برای پیدا کردن الگوهای تکراری استفاده نماییم. به طور مثال: مجموعه‌ای از حداقل تعداد پشتیبانی[18] را برابر با تعداد حداقل خط‌سیرهایی که μ است، قرار می‌دهیم. در اینجا μ=3 قرار داده شده است. و الگوهای {C_1, C_2, C_4 :3} و {C_3, C_5 :3} باید پیدا شوند. پیدا کردن یک مجموعه کامل از تمام الگوهای تکراری ضروری نیست. مجموعه‌ای از ماکزیمم الگوهای تکراری در اطلاعات مورد بررسی بازیابی خواهند شد. مزیت اصلی استفاده از الگوریتم BFEs این است، که الگوهای Flock متوالی‌تری را پیدا می‌کند. مجموعه‌های Closed و Maximal که در الگوریتم BFE-Enhanced وجود دارند، باعث می‌شوند، از الگوهای تکراری پرهیز شود و به یک مجموعه‌ای از پارامترهای δ که بازه‌های زمانی را محدود می‌کنند، نیز نیاز دارند. الگوریتم-BFE Enhanced از δ برای Flockهایی با دوره زمانی مشخص استفاده می‌کند، که الگوریتم BFEs-Enhanced نیز از آن استفاده می‌کند. همچنین الگوهای به دست آمده از شکل شماره ۱ به عنوان خروجی-های معتبری از الگوریتم استخراج الگوهای تکراری به حساب می-آیند.

جدول(۱): نسخه‌ای از تراکنش‌ها در مجموعه داده[۱]

ID تراکنش‌ها	ID دیسک‌ها
T_1	<C_1,C_2, C_3>
T_2	<C_1,C_2,C_4>
T_3	<C_1,C_2,C_4>
T_4	<C_3, C_5>
T_5	<C_3,C_5>
T_6	<C_3,C_5>
T_7	<0>

۳-۳- پیداکردن یک الگوریتم استخراج الگو تکراری در پایگاه داده تولید شده

اولین بار پیدا کردن الگو Flock معرفی شده است در [10,16]، به هرحال آنها، نظریه‌ای درباره دوره‌های زمانی را بررسی نکرده‌اند. برای تشخیص شباهت زیاد Flockها، تنها دو متغیر استفاده شده است. اولین متغیر اشیا متحرک، حداکثر فاصله ثابت و یا Constant Maximum Distance (ϵ) می‌باشد. که این متغیر شعاع دیسک را بیان نموده است. دومین متغیر اشیا متحرک، حداقل تعداد و یا (μ) Minimum Number می‌باشد. هر کدام از این دو متغیر باید در کنار آن دیسک قرار بگیرند. بعد از مدتی، در[۱۹] یک متغیر جدید به نام حداقل دوره زمانی و یا Minimum Duration Time (δ) معرفی شده است که، به عنوان یک پارامتر جدید در Flock مورد بررسی قرار می‌گیرد.

به هر حال در[۱۸]، چگونگی استخراج الگوهای تکراری که باعث می‌شود، یک حجم عظیمی از الگوهای تکراری تولید شود را نشان داده است. البته باید در نظر داشت، که اگر یک الگو تکراری است، تمامی زیرمجموعه‌های آن الگو نیز تکراری است، پس بنابراین، پیچیدگی در تجزیه و تحلیل و درک آن الگوها را به صورت چشمگیری افزایش می‌دهد. به طور کلی، دیدگاه استخراج الگوهای تکراری، یک فرآیندی است که در دهه‌های گذشته برای حل و کاهش الگوهای حرکتی Flock در مجموعه داده‌های خط‌سیر مورد استفاده قرار گرفته است.

شبه کد الگوریتم BFEs-Enhanced به صورت زیر ارائه شده است:

Input: parameters μ, $\in$ و δ
Output: Flocks patterns
$F^{t}_{0} \leftarrow \emptyset$
for each new time instance t_i **do**
  $F^{t}_{i} \leftarrow \emptyset$, C $\leftarrow$ Index, Disks (T[t_i])
  for each c $\in$ C **do**
    for each f $\in F^{t_{i-1}}$ **do**
      if $|c \cap f| \geq \mu$ **then**
        $u \leftarrow c \cap f$
        $u.t_{start} \leftarrow f.t_{start}$
        $u.t_{end} \leftarrow t_i$
        if $(u.t_{end} - u.t_{start}) = \delta$ **then**
          report flock pattern u from $u.t_{start}$ to $u.t_{end}$
          update $u.t_{start}$
        end
        $F^{t}_{i} \leftarrow F^{t}_{i} \cup u$
      end
    end
    $F^{t}_{i} \leftarrow F^{t}_{i} \cup c$
  end
end

۳-۴- ایفاکردن یک روند پس‌پردازش برای چک کردن الگوهای تکراری و سپس هرس کردن الگوهای تکراری به دست آمده

الگوریتم پیشنهادی BFEs-Enhanced مثل سایر الگوریتم‌ها، نیاز به هرس‌کردن الگوهای تکراری و دارای افزونگی دارد. روش‌کار در الگوریتم پیشنهادی همچون الگوریتم BFE-Enhanced است، که برای هرس کردن، یک قطر ثابت برای هر دیسک تعیین می‌کند، این به این دلیل است، که دیسک‌هایی که قطر آنها از قطر ثابت و تعیین شده بیشتر باشد، دارای افزونگی هستند. این در صورتی است که، تمام دیسک‌های دارای افزونگی هم مثل سایر دیسک‌هایی که بدون افزونگی هستند، در بازه‌های زمانی مختلف مشغول به تولید الگوهای Flock هستند، اما دیسک‌هایی که دارای افزونگی هستند، الگوهایFlock آنها هم دارای افزونگی می‌باشد. که در این صورت نیاز به اسکن‌های بیشتری وجود دارد، تا الگوهای Flock دارای افزونگی را در مرحله اول شناسایی، و در مرحله دوم، آنها را حذف نماید. به بیانی دیگر می‌توان، به صورت شکل شماره ۲ به توضیح مراحل هرس‌کردن آن پرداخت.

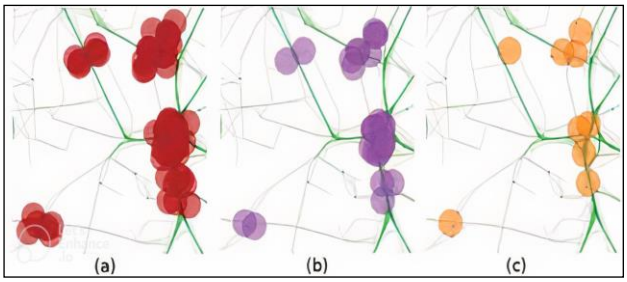
در شکل شماره ۲ یک مجموعه داده از خط‌سیرها و پارامترهایی ($\in$)، (μ) و(δ) به عنوان پارامترهای ورودی تعریف شده است. عملیات کلی در الگوریتم پیشنهادی همچون الگوریتم BFE-Enhanced می‌تواند در سه قسمت به صورت زیر بررسی شود:

۱) توضیح شکل شماره ۲ قسمت(a)، هدف از این شکل، نشان دادن یک مجموع نهایی از یک دیسک است، که برای هر بازه زمانی، حداقل تعداد اشیایی که به یکدیگر نزدیک‌تر هستند، با یکدیگر حرکت خواهند کرد.

۲) توضیح شکل شماره ۲ قسمت (b)، هدف از این شکل، الحاق دیسک‌های نامزد است، که هرکدام همان مجموعه‌ای از اشیایی را که در طول بازه‌های زمانی متوالی اگر و تنها اگر بیشتر از حداقل مقدار μ بودند را به اشتراک می‌گذارند. همچنین پارامتر δ حداقل دوره زمانی را برای اشیا در نظر می‌گیرد.

۳) توضیح شکل شماره ۲ قسمت (c)، هدف از این شکل، با توجه به مسئله Maximal و Closed تشخیص گروه‌هایی از دیسک‌ها که هرکدام به دیسک‌های دیگری تقسیم می‌شوند، و چک کردن آنها که Maximal و یا Closed هستند، نسبت به سایر دیسک‌ها آسانتر می‌شود. که در شکل شماره ۲ محاسبه یک مجموعه معتبر از دیسک‌های نهایی و هرس‌کردن آنها نشان داده شده است.

به طور خلاصه، زمانی که یک مجموعه نهایی از دیسک‌ها برای فواصل زمانی متوالی کشف شده‌اند، سپس الگوریتم پیشنهادی همچون الگوریتم BFE-Enhanced، یکی از دیسک‌ها را در هر مجموعه‌ای که پیدا کرده است را با حداقل تعداد خط‌سیرهای معمولی μ تطبیق می‌دهد. زمانی که یک بازه جدیدی بررسی می‌شود، دیسک‌های جدید با نامزدهای ذخیره شده قبلی الحاق می‌شوند. که در هر لحظه، یکی از آنها نسبت به پارامترδ طولانی‌تر است. و آن سریعا گزارش می‌شود، که باید حذف شود. به هر حال تعداد دیسک‌ها در یک بازه زمانی معین می‌تواند کمی بزرگ و هزینه الحاق آن دیسک‌ها در الگوهای Flock می‌تواند خیلی بالا باشد. همچنین الگوریتم پیشنهادی همچون الگوریتم BFE-Enhanced تعداد نامزدهای ذخیره شده با توجه به دوره زمانی δ را محدود می‌کند. و در نتیجه Flockهایی که در یک دوره زمانی ثابت هستند، را گزارش می‌دهند.



شکل (۲): گام های هرس کردن الگوریتم BFE-Enhanced [1]

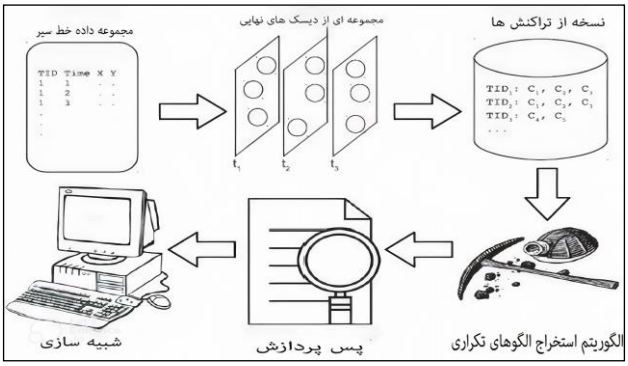
با توجه به شکل شماره۲، عناوین هرکدام از شکل‌ها در شکل شماره ۲ به صورت زیر است:

(a): مجموعه اولیه از دیسک‌ها

(b): تنها دیسک‌هایی که مقدار هرکدام از μ=3 عبور کرده است.

(c): دیسک‌های اضافی با اعضا زیرمجموعه‌ای که حذف شده‌اند.

این چهار مرحله به صورت کامل در فلوچارت روش پیشنهادی در شکل شماره ۳ نشان داده شده است.



شکل (۳): فلوچارت الگوریتم پیشنهادی

۴- یافته‌ها

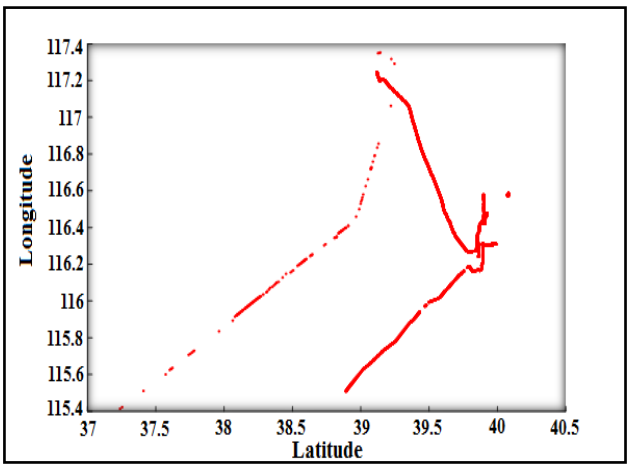
تمامی آزمون‌ها بر روی داده‌های واقعی خط‌سیر به نام پایگاه‌داده Geolife انجام شده است. این مجموعه داده شامل: خط‌سیرهای ۱۸۲ کاربر در طول پنج سال یعنی از آوریل ۲۰۰۷ تا آگوست ۲۰۱۲ می‌باشد. این خط‌سیرها، به صورت دنباله‌هایی از نقاط مکانی در محدوده شهر پکن[۱۹] ثبت شده‌اند، که هر کدام دارای یک مهر زمانی[۲۰] می‌باشند. در پیاده‌سازی‌های مختلف، ما از تعدادی از این داده‌ها بهره‌مند شده‌ایم. که در تشریح هرکدام از الگوریتم‌های استفاده شده، به این مسئله نیز اشاره خواهد شد. تمامی این داده‌ها در سایت شرکت مایکروسافت[۲۱] موجود می‌باشند.

برای پیاده‌سازی قسمتی از داده‌های پایگاه‌داده Geolife با توجه به محدوده‌های زمانی انتخاب شده برای الگوریتم پیشنهادی، که براساس الگوریتم BFE-Enhanced بوده است، از نرم افزار Matlab استفاده گردیده است، همچنین کامپیوتری که بر روی آن پیاده‌سازی انجام گرفته است، دارای مشخصاتی شامل: Operating System = 64, CPU = Intel Core I7, Ram = 8, Windows Version = 8 می‌باشد. در پیاده‌سازی انجام شده بعد از گروه‌بندی مجموعه داده‌های Geolife در داخل فواصل زمانی خاصی، نتایجی به صورت زیر به دست آمد. در اولین پیاده‌سازی که برای پنج‌ماه یعنی از ماه آگوست تا ماه دسامبر ۲۰۰۷ انجام شد، با توجه به جدول شماره ۲ یکسری از معیارهایی در نظر گرفته شد، که مقداردهی این معیارها به صورت جدول شماره ۲ می‌باشد:

جدول (۲): معیارها و مقادیر پیاده‌سازی پنج ماهه

معیار	مقدار
μ	4
Time Threshold	4
Support	30%
Confidence	1

برطبق معیارهای انتخاب شده در جدول شماره ۲، تنها یک الگو Flock در این بازه زمانی پنج ماهه به صورت شکل شماره ۴ تشخیص داده شده است.



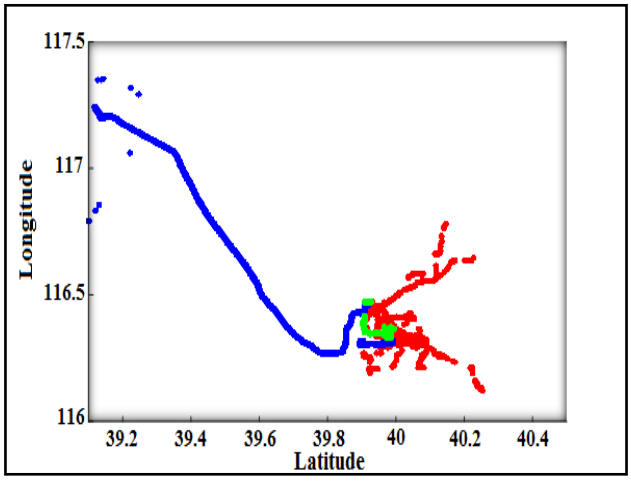
شکل (۴): استخراج یک الگو Flock در بازه زمانی پنج ماهه

در پیاده‌سازی نهایی که برای دوازده‌ماه یعنی از ماه آوریل ۲۰۰۷ تا ماه مارچ ۲۰۰۸ انجام شد، با توجه به جدول شماره ۳ یکسری معیارهایی در نظر گرفته شد، که مقداردهی این معیارها به صورت جدول شماره ۳ می‌باشد:

جدول (۳): معیارها و مقادیر پیاده سازی دوازده ماهه

معیار	مقدار
μ	4
Time Threshold	5
Support	%4
Confidence	1

بر طبق معیارهای انتخاب شده در جدول شماره ۳، تنها سه الگو Flock در این بازه زمانی دوازده‌ماهه به صورت شکل شماره ۵ تشخیص داده شده است.



شکل (۵): استخراج سه الگو Flock در بازه زمانی دوازده ماهه

با توجه به شکل شماره ۵، ما توانستیم سه الگو Flock در بازه زمانی دوازده ماهه به دست بیاوریم، که هرکدام از الگوهای به دست آمده با یک رنگ مشخص شده است. این که هر رنگ مربوط به کدام الگو است در زیر به توضیح آنها خواهیم پرداخت:

- الگو سبز رنگ، یک مجموعه‌ای از خوشه‌های، ماه‌های آوریل، می، ژوئن، جولای، آگوست سال۲۰۰۷ است.
- الگو قرمز رنگ، یک مجموعه‌ای از خوشه‌های، ماه‌های جولای، آگوست، سپتامبر، اکتبر سال ۲۰۰۷ است.
- الگو آبی رنگ، یک مجموعه‌ای از خوشه‌های نوامبر، دسامبر، سال ۲۰۰۷ و ژانویه و فوریه سال ۲۰۰۸ است.

با توجه به توضیحات ذکر شده در بالا، الگوریتم پیشنهادی BFEs-Enhanced مانند: الگوریتم BFE-Enhanced یک متد موثر برای پیدا کردن الگوهای حرکتی Flock می‌باشند. به همین دلیل ابتدا، مجموعه داده‌ها را به پنج بازه زمانی تقسیم نمودیم، یعنی به ازا هر یک ماه، یک بازه زمانی در نظر گرفتیم، که در پیاده‌سازی اولیه پنج بازه زمانی مورد آزمون قرار گرفت. الگوریتم پیشنهادی در تمام این بازه‌های زمانی اعمال شده است. سپس نسخه‌ای از تراکنش‌ها در مجموعه‌ای از داده‌های ایجاد شده است و استخراج الگوهای تکراری در نسخه‌ای از تراکنش‌ها در مجموعه داده‌ها اعمال گردیده است. تمام این موارد گفته شده در مورد آزمون نهایی هم صدق می‌کند، در آزمون دوم، تنها بازهای زمانی به دوازده‌ماه تغیر داده شده است و الگوریتم پیشنهادی همچون آزمون قبلی بر روی آنها اعمال شده است.

نتایج پیاده‌سازی نشان داده است که، الگوریتم پیشنهادی BFEs-Enhanced برای مجموعه داده‌های بزرگ می‌تواند مورد استفاده قرار بگیرد. همچنین این الگوریتم خود توانایی حذف خوشه-های تکراری را دارد، یعنی به حذف خوشه‌هایی با تکرار حداقل می-پردازد. و نیازمند به FP-Tree و یا سایر الگوریتم‌های اینچنینی برای حذف خوشه های تکراری نیست. پس از آزمون‌های انجام شده توسط الگوریتم پیشنهادی BFEs-Enhanced بر روی مجموعه داده‌های پایگاه داده Geolife در مقاطع زمانی مختلف، به مقایسه زمان اجرای الگوریتم پیشنهادی با چهار الگوریتم دیگر که به نام‌های K-Means, Sparrow, SOM, و الگوریتم BFE-Enhanced می‌باشند پرداخته شده است. و نتایج به صورت جدول شماره ۴ نشان داده شده است.

جدول (۴): مقایسه زمان اجرای ۴ الگوریتم

ردیف	تعداد خوشه‌ها	زمان(ثانیه)				
		BFE-E	K-Means	SOM	Sparrow	BFEs-E
1	4	5.37	34.04	13.25	7.32	2.30
2	5	8.51	47.64	22.83	11.19	3.33
3	6	9.40	39.02	38	15.43	3.86
4	7	7.24	33.16	65.32	14.90	4.90
5	8	12.90	48.25	48.30	17.83	6.93
6	9	18.31	57.66	68.30	22.17	11.70
7	10	19.35	79.04	68.30	23.35	16.32

با توجه به جدول شماره ۴ و مقایسه زمان اجرای ۵ الگوریتم BFE-Enhanced، Sparrow، K-means، SOM با الگوریتم پیشنهادی BFEs-Enhanced با استفاده از تعداد خوشه‌های انتخاب شده، این نتیجه حاصل می‌شود که، زمان اجرای الگوریتم BFEs-Enhanced نسبت به چهار الگوریتم دیگر بسیار کمتر بوده است، پس می‌توان گفت که این الگوریتم، الگوریتم بسیار خوبی می‌باشد. زیرا هر چقدر تعداد خوشه ها افزایش یابد، زمان اجرای هر چهار الگوریتم نیز افزایش می‌یابد. اما میزان افزایش زمان اجرای الگوریتم پیشنهادی BFEs-Enhanced بسیار کمتر از میزان زمان افزایش یافته توسط چهار الگوریتم دیگر است.

پس از آزمون نهایی که برای دوازده‌ماه انجام گرفت، نتایج به دست آمده از مقایسه چهار الگوریتم DBSCAN, Sparrow و BFEs-Enhanced برای دوازده‌ماه، که از تاریخ ماه آوریل ۲۰۰۷ تا تاریخ ماه مارچ ۲۰۰۸ می‌باشد. نتیجه به دست آمده از این پیاده‌سازی، سه الگو Flock می باشد. که این نتیجه با آستانه‌های زمانی مختلف، الگوریتم پیشنهادی با سه الگوریتم Sparrow، BFE-Enhanced، DBSCAN در جدول شماره ۵ مقایسه شده است.

جدول (۵): مقایسه تعداد الگوهای حرکتی تشخیص داده شده در چهار الگوریتم

ردیف	آستانه زمانی(μ)	تعداد الگوهای حرکتی تشخیص داده شده			
		BFE-E	DBSCAN	Sparrow	BFEs-E
1	2	7	8	6	5
2	3	6	7	5	4
3	4	5	6	4	3

با توجه به جدول شماره ۵، در این مقایسه با توجه به آستانه زمانی و یا μ که برای هر چهار الگوریتم قرار داده شده است، تعداد الگوهای حرکتی متفاوتی در هر کدام از الگوریتم‌ها به صورت جدول شماره ۵ بدست آمده است. با توجه به نتایج مقایسه انجام شده، الگوریتم BFEs-Enhanced الگوهای کمتری را نسبت به الگوریتم-های DBSCAN, BFE-Enhanced و Sparrow ظاهرا به دست می‌آورد. اما در واقعیت تمام الگوهای بدست آمده توسط آن سه الگوریتم، الگوهای فاقد معنی و مقطعی می‌باشند. به این دلیل که در الگوریتم BFEs-Enhanced معیاری به نام MaxSupport وجود دارد، و این باعث می‌شود، که تمام الگوهای بدست آمده توسط این الگوریتم کاملا متوالی و معنی‌دار باشند.

۵- نتیجه‌گیری

در نهایت، پس از پیاده‌سازی الگوریتم BFEs-Enhanced برای دو بازه زمانی پنج‌ماهه و دوازده‌ماهه به ترتیب یک و سه الگو Flock استخراج شد. سپس به مقایسه الگوریتم BFEs-Enhanced با چهار الگوریتم دیگر پرداخته شد، که با توجه به نتایج به دست آمده از مقایسه، الگوریتم BFEs-Enhanced دارای زمان اجرای پایین‌تر نسبت به سایر الگوریتم‌ها می‌باشد، همچنین تعداد الگوهای استخراج شده توسط این الگوریتم نسبت به سایر الگوریتم‌ها کمتر است، اما تمام الگوهای استخراج شده توسط این الگوریتم به صورت متوالی و معنی‌دار می‌باشند. همچنین در این مقاله، با استفاده از روش پیشنهادی، معیار MaxSupport را به سایر معیارهای موجود در

الگوریتم BFE-Enhanced اضافه نمودیم، و الگوریتم پیشنهادی BFEs-Enhanced را ارائه نمودیم، و با بالابردن مقدار معیار MaxSupport الگوهای متوالی‌تر و معنی‌دارتری و شاید کمتری را از الگوریتم پیشنهادی استخراج نمودیم. همچنین در الگوریتم پیشنهادی برای خوشه‌بندی مسیرها و بالا بردن میزان دقت خوشه-بندی از دو معیار فاصله و جهت بهره برده شده است.

باتوجه به این که این حوزه، بسیار گسترده و جدید محسوب می‌شود، پیشنهادات زیادی را می‌توان برای سایر پژوهش‌ها در آینده داشت، از جمله پیاده‌سازی الگوریتم Traclass بر روی پایگاه‌داده‌های Geoife و مقایسه نتایج این الگوریتم‌ها با نتایج به دست آمده از الگوریتم BFEs-Enhanced را نیز می توان مورد بررسی قرار داد.

مراجع

[1] A. O. Calderon Romero "Mining Moving Flock Patterns Using Spatio-Temporal Datasets." in Computer and Electronic Department, Master of Science (MSc.) University of Southampton (UK). 2011.

[2] T. Zhang, W. He, J. Huang, Z. He, J. Li "Interactive Visual Analytics of Moving Passenger Flocks Using Massive Smart Card Data" Presentation in International Journal of Taylor& Francis Vol. 49, Issue. 4, March 20 22, pp. 354-369. DOI:10.1080/15230406.2022.2039775.

[3] M. Erfelic, T. Caric, T. Erdelic, L. Tisljaric. "Transition State Matrices Approach for Trajectory Segmentation Based On Transport Mode Change Criteria." Presentation in International Journal of Sustainability (MDPI). Vol. 14, Issue. 5, February 2022, PP. 2-20. DOI: 10.3390/14052756.

[4] Z. Shi and L.S.C. Pun-Cheng. "Spatiotemporal Data Clustering: A Survey of Methods." Presentation in International Journal of Geo-Information Vol. 8, No.112, February 2019, PP. 2-16. DOI: 10.3390/ijgi8030112.

[5] G. Yuan, P. Sun, J. Zhao, D. Li and C. Wang. "A Review of Moving Object Trajectory Clustering Algorithms." Presentation in Artificial Intelligence Review an International Science and Engineering Journal Springer Vol. 47, No.1, March 2016, PP. 1078-1090. DOI: 10.1007/s10462-016-9477

[6] C. Aggarwal, M. Bhuiyan and M. Hasan. "Frequent Pattern Mining." Presentation in Frequent Pattern Mining Algorithms: A Survey Springer, July 2014, PP.19-28. DOI: 10.1007/978-3-319-07821-2_2

[7] A. Boulmakoul, L. Karim, and A. Lbath. "Moving Object Trajectories Meta-Model and Spatio -Temporal Queries." Presentation in International Journal of Database System (IJDMS) Vol.4, No.2, April 2012, PP. 35-54. DOI: 10.5121/ijdms.2012. 4203.

[8] H. Alberto García Salas, A. Gelbukh, H. Calvo and F Galindo Soria. "Automatic music composition with simple probabilistic generative grammars", Presentation in International Journal of Polibits, Vol.44, No. 44, July 2011, PP. 59-65.

[9] G. Folinio, A. Forestiero, and G. Spezzano. "An Adaptive Flocking Algorithm for Performing Approximate Clustering." Presentation in International Journal of Information Science in Elsevier Vol. 179, No.18, Jan 2010, pp.3059-3078. DOI: 10. 1016/j.ins.2009.05.017

[10] M. Kwan Shan and L-Y. Wei. "Algorithms for Discovery of Spatio-Orientation Patterns from Images." Presentation in International Journal Expert Systems with Application in Science Direct. Vol. 37, No.8, August 2010, pp. 5795-5802, DOI: 10.1016/j.eswa. 2010.02.028

[11] M. Wachowicz, C. Renso and M. Nanni. "Finding Moving Flock Among Pedestrians Through Collective Coherence." Presentation in International Journal of Geographical Information Science. Vol. 25, No.11, February 2010, pp.1849-1864. DOI: 10.1080/13658816 .2011.561209.

[12] J. Kang, H. Ma, Z.Duan, H. He. "Vehicle Trajectory Clustering in Urban Road Network Environment Based on Doc2Vec Model." Presentation in International Joint Conference on Neural Networks (IJCNN) in IEEE. China, 18-22 July 2021, DOI: 10.1109/IJCNN52387.20 21.9534341.

[13] Y. Cao, J. Zhu, F. GAO"An Algorithm for Mining Moving Flock Patterns from Pedestrian Trajectories" Presentation in Conference on Asia-Pacific Web Conference, Japan, September 2016. DOI: 10.1007/978-3-319-45835-9_27.

[14] C. Zhang, J. Han, L. Shou, J. Lu and T. La Porta. "Mining Fine grained Sequential Patterns in Semantic Trajectories." Presentation in International Conference on Very Large Data Bases. Hangzhou, China, 5 May2014. DOI: org/10.14778/2732939.2732949.

[15] K. Khan, S. Ur Rehman, K. Aziz, S. Fong and S. Sarasvady. "DBSCAN: Past, Present and Future." Presentation in the Fifth International Conference IEEE on the Applications of Digital Information and Web Technologies (ICADIWT). Bangalore, India, 15 May 2014. DOI: 10.1109/ICADIWT.2014.6814687.

[16] B. Hu and M. Ester. "Spatial Topic Modeling in Online Social Media for Location Recommendation." Presentation in 7th ACM Conference on Recommender System. Hong Kong, China, 12 October 2013. DOI:.org/10.1145/2507157.2507174

[17] G. Miriam Jacob and S. Mary Idicula. "Detection of Flock Movement in Spatio-Temporal Database using Clustering techniques- An Experience." Presentation in IEEE International Conference on Data Science & Engineering (ICDSE), Cochin, Kerala, India, 18-20 July 2012. DOI: 10.1109/ICDSE.2012.6282312.

[18] L. Hong, A. Ahmed, S. Gurumurthy and A. J. Smola. "Discovering Geographical Topics in the Twitter Streem." Presentation in ACM International World Wide Web Conference Committee (IW3C2). Lyon, France, 2 April 2012. DOI: 10.1145/2187836.2187940.

[19] Z. Yin, L. Cao, J. Han, C. Zhai and T. Huang. "Geographical Topic Discovery and Comparison." Presentation in ACM International World Wide Web Conference Com-mittee (IW3C2). Hyderabad, India, 28 March2011.DOI:https://doi.org/10.1145/1963405.196344 3.

[20] Z. Yin, L. Cao, J. Han and J. Luo. "Diversified Trajectory Pattern Ranking in Geo-Tagged Social Media." Presentation in Proceeding of the Eleventh SIAM Conference on Data Mining. Mesa, Arizona, USA. 28-30 April 2011. DOI: 10.1137/1.9781611972818.84.

[21] Z. Chen, H. Tao Shen and X. Zhou. "Discovering Popular Routes from Trajectories", Presentation in IEEE 27th International Conference on Data Engineering ICDE, Hannover, Germany, 16 April 2011. DOI: 10.1109/ICDE.2011.5767890

[22] H. Htet Aung and K-L. Tan. "Discovery of Evolving Convoys." Presentation in Springer International Conference on Scientific and Statistical Database

Management. Berlin, Heidelberg, 19-21 Jun 2010. DOI https://doi. org/10.1007/978-3-642-13818-8_16.

[23] R. Singh, T. Johnsten, V. Raghavan and Y. Xie. "Efficient Algorithm for Discovering Potential Interesting Patterns with Closed Item sets." presentation in IEEE2010 IEEE International Conference on Granular Computing .San Jose Jose, CA, USA, 14-16 Aguste 2010. DOI: 10.1109/GrC.2010.55.

[24] T. Kurashima, T. Iwata, G. Irie and K. Fujimura. "Travel Route Recommendation using Geotags in Photo Sharing Sites." Presentation in ACM International Conference on Information and Knowledge Management. Torento-Canada, 26 October 2010. DOI: 10.1145/1871437.187 1513.

[25] M. R. Vieira, P. Bakaloy and V. J. Tsotras "on line Discovery of Flock Pattern in Spatio-Temporal Data" Presentation in ACM International Conference on Advances in Geographic Information Systems. Seattle, Washington 4-6 November, 2009. DOI: https://dl.acm. Org/citation.cfm? Id =1653812.

[26] H. Jeung, H. Tao Shen and X. Zhou. "Convoy Queries in Spatio-Temporal Databases." Presentation in IEEE 24th International Conference on Data Engineering. Cancun, Mexico, 7-12 April 2008. DOI: 10.1109/ICDE.2008.44 97588.

[27] P. Petrou, P. Tampakis, H. Georgion, N. Pelekis, Y. Theodoridis. "Online Long-Term Trajectory Prediction Based on Mined Route Patterns." Presentation in International Workshop on Multiple-Aspect analysis of Semantic Trajectories with Springer, 4 January 2020, DOI: 10.1007/978-3-030-38081-6_4.

زیرنویس‌ها

[1] Trajectory
[2] Moving Objects
[3] Time-Space
[4] Datamining
[5] Trajectory Pattern Mining
[6] Trajectory Clustering
[7] Location Base Semantic
[8] Moving Cluster
[9] Maximum Distance
[10] Candidate Pattern
[11] Corresponding
[12] Frequent Item set Mining Implementation
[13] Constant Maximum Distance
[14] Minimum Number
[15] Minimum Duration Time
[16] Sequential
[17] Life Time
[18] Minimum Support(min_sup)
[19] Beijing
[20] Time Stamp
[21] www.Microsoft.com