به نام خداوند بخشنده و مهربان

## مقاله: "Show me how you move and I will tell you who you are"

1. به طور خاص، هدف از یک حمله استنتاج به داده های موقعیت جغرافیایی چه میتواند باشد؟

یک مهاجم به داده های موقعیت جغرافیایی ممکن است اهداف گوناگونی داشته باشد، ولی به طور دقیق تر میتوان اهداف را به صورت زیر دسته بندی کرد:

- بدست آوردن مکانهای مهم(POI[1] ها) برای دستیابی به علایق شخصی
- پیش بینی الگوی حرکتی فردی اعم از مکان گذشته، حال و آینده شخص
- یادگیری معنادار رفتار حرکتی شخصی از روی POIها و الگوی حرکتی
- پیوند بین اطلاعات شخصی افراد
- کشف روابط اجتماعی بین فردی

2. چند نمونه از الگوریتم های یادگیری و روش های استنتاج را توضیح دهید.

- *خوشه بندی (Clustering)* یکی از روش های بدون نظارت یادگیری است که تلاش میکند که اشیای مشابه را در یک خوشه و اشیای نامشابه را در خوشه های متفاوت قرار دهد.
- *مدل حرکتی* از روی اطلاعات موقعیت جغرافیایی یادگرفته میشود و سپس برای شناسایی آنها از بین مجموعه داده های موقعیت جغرافیایی و یا پیش بینی حرکات بعدی استفاده میشود.
- *ابتکارات (Heuristics)* نیز در عمل نتایج خوبی برای شناسایی POI ها با هزینه نسبتا پایین ارائه میکند.
- *داده هایی که از برنامه های کاربردی اجتماعی* مانند توئیتر کسب میشود، یکی از منابع ممکن اطلاعاتی دشمن است که نقشه حمله به حریم خصوصی فردی را طراحی کند.
- *اطلاعاتی که از منابع عمومی* مانند نقشه گوگل بدست می آید، یک منبع بالقوه از دانش است که دشمن از آنها بهره برداری میکند.

3. مکانیزم های بازسازی جغرافیایی چیست؟ چرا از این مکانیزم ها استفاده میشود؟

[1] Points of Interests

الگوریتم بازسازیS ، یک مجموعه داده های موقعیت جغرافیاییD میگیرد و برخی عدم اطمینان به داده ها اضافه میکند و بعضی از اطلاعات را برای افزایش حریم خصوصی افراد، از D حذف میکند و مجموعه داده جدید D′ را تحویل میدهد. ایده اصلی بازسازی جغرافیایی این است که برای یک دشمن بالقوه نقض حریم خصوصی یک کاربر خاص با استفاده از D′ سخت تر از D باشد است.

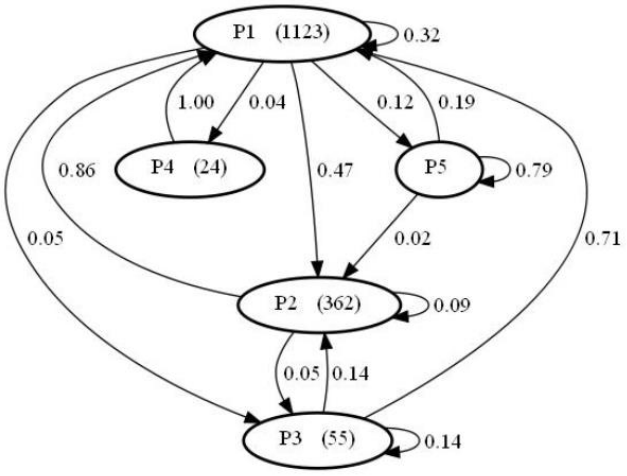
4. برای بازسازی جغرافیایی از چه تکنیکهایی استفاده میشود؟

- استفاده از *اسم مستعار* بجای اسم مشخص و رایج آن، یا استفاده از اسم ناشناخته.
- *روشهای اختلال*، فضای مختصات متحرک را با اضافه کردن برخی اغتشاش تصادفی تغییر میدهد.
- *تجمع*، ردپای چند متحرک را به یک مختصات یکسان ادغام میکند.
- *نمونه برداری* میتواند به عنوان یک فرم از تجمع زمانی به کار برده شود.
- *پنهان کردن فضایی*، یک شکل از تجمع از حوزه فضایی-زمانی است.
- *مخلوط مناطق* که از مخلوط شبکه ها الهام گرفته شده است.
- *جابجایی و مبادله*، شامل جابجا کردن آثار تحرک دو فرد متفاوت برای دوره زمانی خاص میشود.
- *پاک کردن و از بین بردن* ردپای حرکتهایی که خیلی حساس تلقی میشوند، یک روش پاکسازی است.

5. زنجیره تحرک مارکوف را توضیح دهید.

زنجیره حرکت مارکوف مدلی است برای نمایش فشرده و نسبتا دقیق رفتار حرکتی یک فرد. در واقع زنجیره مارکوف یک ماشین احتمالاتی است که در آن هر استیت آن نشان دهنده POI ها و احتمال ها، نشان دهنده انتقال بین POI هاست. به عبارت دیگر میتوان گفت که زنجیره مارکوف حرکت سیستم انتقال است که از یک مجموعه استیت ها ($P = \{p_1, \dots, p_n\}$) و یک مجموعه انتقال ها ($T = \{t_{1,1}, \dots t_{n,n}\}$) تشکیل شده است.

6. چه اطلاعاتی از زنجیره تحرک مارکوف زیر میتوان استخراج نمود؟ توضیح دهید.



با نگاه کردن به نمودار بالا میتوان دید که استیت P1 از همه استیت ها با احتمال بالایی قابل دستیابی است. همچنین وزن این استیت به مراتب بیشتر از بقیه است، پس میتوان گفت که استیتP1 "منزل" کاربر می باشد. در ادامه اگر بخواهیم مکان محل کار کاربر را پیدا کنیم با استفاده از این گزاره که «مردم غالباً از منزل به محل کار میروند و برمیگردند» میتوان گفت P2 به احتمال زیاد "محل کار" کاربر است. برای استیت P3 میبینیم که از منزل و محل کار به آن دسترسی داریم و احتمالات گویای این است که کاربر از آنجا به منزل میرود. طبق گزاره ی «در عصر، مردم معمولاً بعد از کار قبل از برگشتن به خانه ورزش میکنند.» میتوان فهمید که P3 محل ورزش کاربر است. در انتها میتوان دید که استیت P4 تنها با خانه در ارتباط میباشد، این استیت بهترین گزینه برای کارهای آخر هفته مانند مکان تفریح و فراغت یا محل خرید باشد. و استیت P5 هم یکی از "POIهای نادر" محسوب میشود.

7. GEPETO[2] چیست؟ کاربرد و معماری آن را توضیح دهید.

هدف اصلی GEPETO این است که به محققان مرتبط با حریم خصوصی جغرافیایی وسیله ای به منظور بررسی روشهای پاکسازی مختلف و حملات استنتاج بر روی داده های موقعیت جغرافیایی ارائه کند. ایده اصلی ارائه یک ابزار عمومی و انعطاف پذیر است.

GEPETO با معماری چند لایه طراحی شده با هدف ایجاد سیستم کاربردی، کارآمد، مقیاس پذیر، به راحتی قابل تغییر و قابل اعتماد است. لایه های آن عبارتند از: لایه داده، لایه کنترل، لایه کاربردی و لایه تجسمی.

هدف از این معماری لایه ای، جدا کردن دسترسی به داده و نمایش آن است. که این باعث میشود به راحتی الگوریتم های جدید در لایه کاربردی پیاده سازی شود، بدون نگرانی در مورد چگونگی دسترسی لایه های کنترل و ارائه به داده های تجسمی.

8. چه الگوریتم هایی برای آزمایش های این مقاله در GEPETO مورد استفاده قرار گرفته است؟

- الگوریتم خوشه ای تراکم-پیوستن (Density-Joinable cluster / DJ)
- الگوریتم خوشه ای تراکم-زمان (Density-Time cluster / DT)
- الگوریتم خوشه ای زمان-تراکم (Time-Density cluster / TD)

9. کارآمدی حمله استنتاج به چه عواملی بستگی بیشتری دارد؟ الگوریتم های مورد استفاده را مقایسه کنید.

کارآمدی حمله استنتاج شدیدا به فرایند بهسازی داده های هدف وابسته است. به عنوان مثال در روش استفاده از نمونه برداری برای بهسازی داده ها، الگوریتم خوشه ای DT یک فراخوان بالا و دقت خوب را ارائه میدهد، ولی عملکرد آن در پاسخگویی به اختلال و به طور قابل توجهی کاهش میابد. پس اگر تنها از روش نمونه برداری برای بهسازی داده ها استفاده میکنیم، دشمن تنها با استفاده از الگوریتم خوشه ای DT یک حمله کارآمد داشته باشد. از طرفی الگوریتم TD جایگزین مناسبی برای هر دو روش نمونه برداری و اختلال میباشد که عملکرد آن تحت این دو نوع آشفتگی خوب باقی میماند.

10. چه روشهایی برای حفاظت از امنیت جغرافیایی وجود دارد؟

[2] GEoPrivacy Enhancing TOolkit

در این مقاله اساسا از روش بهسازی داده استفاده شده است. ولی روشهای دیگری از جمله رمز کردن اطلاعات اولیه، مکانیزم های کنترل دسترسی برای موجودیت خارجی نیز برای این منظور وجود دارد. در بعضی جهات این دو رویکرد مکمل روش بهسازی داده میباشند.