Identifying and Analyzing Patterns of Evasion (HM0210-13-1-0005)
Shashi Shekhar, University of Minnesota, shekhar@umn.edu
Apr 29, 2019

Abstract
We developed space-time aware methods for modeling patterns of evasive behavior by insurgents and other security targets. These targets increase employ techniques to mask their movements and locations. Denial, deception and evasion techniques skew data collection and hinder traditional data mining techniques. We proposed an overarching framework to identify and analyze these denial and deception instances. First, we proposed a method to distinguish between evasive and non-evasive behaving groups by quantifying the space-time entropy or predictability of individuals’ behavior. Second, we identified “blackholes” areas where no target movement is observed, despite predictions that such movement would occur. Third, since conventional data mining techniques cannot be applied in areas lacking reported observations (blackholes), we looked to theoretical understanding of human behavior to help generate hypotheses about target location and travel routes. Specifically, we applied routine activity theory, a well-known theory used by environmental criminologists, which holds that individuals typically follow set patterns in their daily lives.

Key Words: Evasive pattern mining, blackhole pattern, environmental criminology

This report follows the format guidance from the contract.
1.0 Executive Summary

This project explored an overarching framework to identify and analyze certain denial and deception instances by providing a taxonomy of evasive patterns based on data types such as points and trajectory. Then, we explore algorithms for areas where no target movement is observed, despite predictions that such movement would occur. Finally, we looked to theoretical understanding of human behavior to help generate hypotheses about target location and travel routes.

**Background:** Related data mining work (e.g., hotspot and anomaly detection for movement analysis) identify clusters of objects (moving or static), increased activity on road networks, and other directly observable phenomena. However, the related work assumes that there is no denial of data and is thus not effective in identifying evasive patterns. To overcome the limitations of related work, our approach provides the concepts of blackholes and avoidance patterns around the lack of observable data points.

**Technical approach:** We classified evasive patterns based on the input data type (e.g., points, trajectories) and model evasive behaviors in point and trajectory data as blackhole and avoidance patterns, respectively. Further, we describe algorithms for Ring-Shaped Hotspot Detection (RHD), Avoidance Region Miner (ARM) etc.

**Finding/Results:** Ring-shaped hotspot detection technique was effective in localizing the source of Legionnaire disease from point dataset from Manhattan, New York. Avoidance Region Miner technique was effective in finding avoidance zones in an open street map trajectory dataset from Rome, Italy.

**Analysis, products, and discussion:** Case studies, experiments and analytical evaluation showed that the new algorithms for Ring-Shaped Hotspot Detection (RHD), Avoidance Region Miner (ARM) improved computational scalability without impacting result quality.

Future work includes consideration of temporal information while identifying evasive patterns, using holistic trajectory data for monitoring evasive maritime behaviour etc.
2.0 Introduction

2.1 A comparison of actual accomplishments with the goals and objectives established, the findings of the investigator, or both.

In this project, we developed an overarching framework to identify and analyze instances of denial and deception by enemy insurgents. Our first goal was to develop a means to distinguish between evasive and non-evasive behaving groups. Our second goal was to be able to identify “blackholes,” areas where no target movement is observed, despite predictions that such movement would occur. Third, since conventional data mining techniques cannot be applied in areas lacking reported observations, we looked to theoretical understanding of human behavior to help generate hypotheses about target location and travel routes.

Evasive vs. non-evasive behaving groups: Our first goal was to develop a method to model the difference between evasive and non-evasive behavior and classify movement patterns.

Key Accomplishments: We developed a taxonomy to define and classify evasive patterns based on the input data types, including point data, trajectory data, and others (e.g., satellite imagery).

“Blackholes” and patterns of evasion: The goal of this phase of the research was to formally define “blackhole” areas and evasive patterns and develop computational methods to mine them from real-world movement datasets.

Key Accomplishments: We modeled “blackholes” as evasive Ring-shaped Hotspot patterns based on environmental criminology theory, which says that criminals tend to commit crimes neither too close nor too far from their home. We formally defined the problem of Ring-Shaped Hotspot Detection (RHD) to identify ring-shaped hotspots and developed statistical models and computational approach to solve RHD from large scale datasets. Results were published in IEEE Transactions on Knowledge and Data Engineering in 2016.

Theory-based movement hypothesis generation for target location and travel routes: The goal of this research was to develop a model for generating hypotheses about human movement and travel route selection which reflects real human evasive behaviors.

Key Accomplishments: We generated a hypothesis about avoidance behavior in human movement based on theories in anthropology about energy minimization in route selection with consideration of realistic scenarios such as characteristics of commute and exploration routes. Based on the hypothesis, we developed an Avoidance Region Miner (ARM) that identifies avoidance regions from movement datasets. ARM was published in SIAM International Conference on Data Mining in 2018.
2.2 Participants and contributions

Participants included Prof. Shashi Shekhar, and graduate students Emre Eftelioglu, Xun Tang and Arun Sharma (all University of Minnesota) along with Dr. Christopher Farah (NGA Innovision). Prof. Shekhar provided the overall management and guided the graduate students. Emre Eftelioglu investigated on Ring Shaped Hotspot Detection. Emre and Xun Tang probed the Avoidance Patterns. Arun Sharma searched for new trajectory dataset for maritime ship trajectories. In addition, Dr. Christopher Farah served as the NGA technical point of contact and reviewed research ideas as well as technical slides and papers.

3.0 Background

Finding evasive patterns are important in sociology, city/transportation planning and crime analysis. Insurgents and other security targets employ techniques to mask their movements and locations through denial, deception and evasion techniques which skew data collection and hinder traditional data mining techniques. Entropy is a measure of unpredictability, so a space-time entropy measure applied to a target would quantify the unpredictability of a target’s movement. We researched on entropy measures for use in identifying target groups based on historical movement patterns. Although Shannon’s Diversity Index (SDI) is a commonly used measure of entropy, there is a substantial semantic gap between SDI and the notion of spatio-temporal predictability. Closing this semantic gap requires conceptual and computational innovations. Therefore, we proposed to develop a conceptual model to represent the spatio-temporal predictability of targets’ movement across a space.

Blackholes are areas where the denial, deception and evasion techniques of target movements occurs. Related work in hotspot and anomaly detection for movement analysis identifies clusters of objects (moving or static), increased activity on road networks, and other directly observable phenomena. The approach we developed makes it possible to consider the lack of observable data points, or essentially, to identify the opposite of a hotspot.

Due to the lack of target observations inside identified blackholes, it is difficult to use traditional data mining methods to generate hypotheses regarding target movement. Once blackhole areas have been identified, however, routes and locations of interest can be hypothesized based on previously recorded observations using existing theories in relevant fields (e.g., routine activity theory in environmental criminology, kinematics in physics), the underlying transportation network and mobility expectations such as continuity of trajectories over space-time.

Routine activity theory and environmental criminology states that humans tend to follow personal schedules within their own awareness spaces (e.g., favorite places). Modeling these routine activities may allow for new knowledge to be mined from movement datasets, potentially highlighting anomalous activities such as visiting new places (anomalous places) or the visiting of known places at uncommon times (anomalous visits).

This report follows the format guidance from the contract.
One interesting pattern that criminologists have studied, known as the doughnut-hole pattern (Brantingham and Brantingham, 1981), can be considered a representation of a blackhole. Doughnut-hole patterns are formed because of the constant tension between an offender’s desire to divert attention from his or her home base and the desire to travel no further than necessary to commit a crime. We conceptualize doughnut holes as ring shaped hotspots, a special type of concentration of activities.

4.0 Technical Approach

First, we created a taxonomy for classifying evasive patterns based on the input data types (shown in Figure 1) and we modeled evasive behaviors in point and trajectory data as blackhole and avoidance patterns, respectively. We also developed several novel algorithms to identify these patterns from large-scale datasets.

![Figure 1. A taxonomy of evasive patterns](image)

For point data, based on theories in environmental criminology, we formalized “blackhole” patterns with the notion of Ring-Shaped Hotspot (RHD) detection using concentric rings (in which the inner and outer circle of a ring share the same center) where the concentration of activities inside the ring is much higher than outside. RHD is challenging due to the large number of candidate rings generated and high computational cost of the statistical significance test. To address these challenges, we first proposed a novel approach named enhanced dual grid based prune and refine (DGP) algorithm which features a dual grid based upper bound likelihood pruning approach which removes redundant param-grid cells from the result set in pruning phase that improves computational scalability without impacting result quality. We then implemented several new techniques for improving the performance of DGP: a new Prune Phase Algorithm with Multi Cell Size Pruning and Local Maxima Elimination and a new Refine Phase Algorithm with Best Enclosing Ring Approach.

For trajectory data, we formally defined an avoidance region pattern based on movement hypothesis generation methods for avoidance behaviors developed based on theories in anthropology about energy minimization in route selection with consideration of realistic scenarios such as characteristics of commute and exploration routes and formulate avoidance regions. In

This report follows the format guidance from the contract.
order to identify avoidance regions, we developed an “Avoidance Region Miner (ARM)” where we avoid unnecessary candidate avoidance region enumeration by initially creating the avoidance polygons from each trajectory and shortest path pair, then merging these to create a larger avoidance polygon and finally getting an avoidance sub-graph of the spatial network which improves the scalability without losing any patterns present in the data. We conducted both experimental and theoretical analysis to evaluate the performance of proposed approaches in both qualitative and quantitative manner.

5.0 Findings/Results

*Ring-shaped hotspot detection (RHD):* Experimental and theoretical evaluations of our proposed approach for RHD, Dual Grid Based Pruning with Local maxima elimination, Multi cell length and Refine with best enclosing ring (DGPLLMR), demonstrated that **DGPLLMR** yields substantial computational savings compared to baseline approaches. A case study using a real-world disease outbreak dataset (shown in Figure 2) showed **DGPLLMR** is able to find new significant RHD patterns that the widely used SaTScan program misses.

![Figure 2: Input Legionnaire cases (left), Results from SaTScan (middle), Results from RHD (right)](image)

*Avoidance Region Miner (ARM):* Experimental and theoretical evaluations showed that ARM, with two novel algorithmic refinements, yields substantial computational savings without reducing result quality. In a case study, whose results are shown in Figure 3, ARM was able to find novel avoidance region patterns using a real-world trajectory dataset in Rome, Italy.
This report follows the format guidance from the contract.

6.0 Analysis, Products, and Discussion

*Ring-shaped hotspot detection (RHD):* For identifying RHD defined on point data, we proposed a Dual Grid Based Prune with Local maxima elimination, Multi cell length and Refine with best enclosing ring (DGPLMR) algorithm which features a dual grid based upper bound likelihood pruning approach with Multi Cell Size Pruning and Local Maxima Elimination and a refined phase algorithm with Best Enclosing Ring Approach. DGPLMR improves computational scalability without impacting result quality and is validated via case studies, experiments and analytical evaluation. Figure 2 shows Legionnaire disease from point dataset from Manhattan, New York where street where SaTScan returned two small circular hotspots, while RHD returned a single ring-shaped hotspot with a low p-value (i.e., p-value 0.001) and a high Log Likelihood Ratio. The results were published in the IEEE Transactions on Knowledge and Data Engineering journal in 2016 and IEEE International Conference on Data Mining in 2014. The code is being summarized into Github.

*Avoidance Region Miner (ARM):* For identify avoidance regions defined on trajectory data, we proposed an algorithm namely “Avoidance Region Miner (ARM)” which improve the scalability of a naïve approach without losing any patterns present in the data. ARM with the proposed algorithmic refinements is evaluated qualitatively via case studies and quantitatively via experimental and theoretical analysis. Figure 3 shows open street map trajectory dataset from Rome, Italy where left figure shows real dataset of 1312 vehicle GPS trajectories and right figure shows interesting avoidance regions on the north are around the British Ambassador’s residence as well as two hospitals where drivers may be avoiding this area due to closed streets or congestion caused by increased security measures. The results have been published in the proceedings of SIAM International Conference on Data Mining in 2018. The code is being summarized into Github.
7.0 Conclusion

In this project, we have developed multiple space-time aware methods for modeling patterns of evasive behavior by insurgents and other security targets based on the input data types such as point and trajectory data. For point data, we formulated the problem of Ring-shaped Hotspot Detection (RHD) based on theories in environmental criminology and developed the novel DGPLMR algorithm to solve RHD. For trajectory data, we developed methods for evasive behavior hypothesis generation and defined avoidance region patterns based on theories in anthropology about energy minimization in route selection with consideration of realistic scenarios such as characteristics of commute and exploration routes. To detect avoidance regions, a novel algorithm named Avoidance Region Miner (ARM) was proposed. The performance of the models and approaches in the project have been demonstrated via both theoretical and experimental analysis including case studies using real-world datasets.

8.0 Future Work

In future, we plan to extend our current work towards the following directions. First, we plan to consider the temporal information in identifying evasive patterns from point data which could help understand the evolution of pattern. In addition, we will investigate other models for avoidance zone which distinguish detours made because of avoiding a region or other reasons such as exploring some nearby regions. Moreover, we plan to consider evasive patterns using holistic trajectory data which have attributes associated with each point. For example, ship trajectories recorded by automatic identification system (AIS) contain many attributes (e.g., draught) which potentially provide critical information for maritime evasive behaviors such as illicit fishing and dumping activities. Furthermore, we will develop models and approaches for discovering evasive pattern using other data sources (e.g., satellite imagery).
References

Books


Journal Articles


This report follows the format guidance from the contract.
Tab 1: Publications


Tab 2: Student Supported

<table>
<thead>
<tr>
<th>Name</th>
<th>Graduate/ Undergraduate</th>
<th>Degree Awarded</th>
<th>Thesis/ Dissertation Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>Emre Eftelioglu</td>
<td>Graduate</td>
<td>Ph.D. in Computer Science</td>
<td>Geospatial Data Science to Identify Patterns of Evasion</td>
</tr>
<tr>
<td>Xun Tang</td>
<td>Graduate</td>
<td>Ph.D. in Computer Science (in progress)</td>
<td></td>
</tr>
<tr>
<td>Arun Sharma</td>
<td>Graduate</td>
<td>Ph.D. in Computer Science (in progress)</td>
<td></td>
</tr>
</tbody>
</table>
### Tab 3: Algorithms/Code Developed

<table>
<thead>
<tr>
<th>Name</th>
<th>Algorithm/Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dual grid based prune and refine (DGP) algorithm</td>
<td>Algorithm and Source Code</td>
<td>DGP discovers ring-shaped hotspots from a collection of geo-referenced events. It features a dual grid based upper bound likelihood pruning approach that improves computational scalability without impacting result quality</td>
</tr>
<tr>
<td>Dual Grid Based Prune with Local maxima elimination, Multi cell length and Refine (DGPLMR) algorithm</td>
<td>Algorithm and Source Code</td>
<td>DGPLMR improves the performance of DGP by introducing multiple new algorithmic refinements including (1) a new Prune Phase Algorithm with Multi Cell Size Pruning and Local Maxima Elimination and (2) a new Refine Phase Algorithm with Best Enclosing Ring Approach.</td>
</tr>
<tr>
<td>Avoidance Region Miner (ARM) algorithm</td>
<td>Algorithm and Source Code</td>
<td>ARM identifies avoidance patterns from GPS trajectory datasets. It yields substantial computational savings by two algorithmic refinements: (1) elimination of redundant candidate avoidance regions enumeration, and (2) Elimination of Unnecessary Interestingness Ratio Computation.</td>
</tr>
</tbody>
</table>

### Tab 4: Transitions

None.

This report follows the format guidance from the contract.