Predicting a Serial Criminal's Next Crime Location Using Geographical Profiling
利用地理学分析预测连环罪犯的下一个犯罪地点

Bryan Ward
Ryan Ward
Dan Cavallaro

COMAP Mathematical Contest in Modeling
February 22, 2010
Bucknell University

翻译：周吕文

关注“数学模型”公众号，回复“论文”，获取更多美赛论文和翻译

[0]若发现翻译问题，请邮件告知，谢谢。

微信公众号·数学模型 MATHmodels

Abstract

Geographical profiling techniques can be useful to law enforcement agencies who are investigating a serial criminal. We develop such methods that determine the probable location of a serial criminal's next crime based on the spatiotemporal data of their previous crimes. We consider standard deviation, centralization, and probability distance methods for prioritizing a given search area, and also develop analogous methods which weight the spatial data of recent crimes more heavily. We then develop ways of aggregating results from multiple methods. The performance of a geographical profiling method is based on its effectiveness in narrowing down a particular search area into regions likely to contain the next crime location. All of our methods produce a prioritized search area in which the next crime occurs in approximately the top 10% of the search area, a significant improvement over a uniform random distribution. However, the differences in performance among the different methods developed were statistically insignificant. We also found the accuracy of our methods varied depending on the serial criminal under investigation.

对于正在调查某一连环犯罪 (或系列犯罪) 的警方 (执法机构) 而言, 地理学分析技术是非常有用的. 根据连环犯罪以往的作案时间和地点, 本文建立了预测下一次作案可能的发生地点的地理分析算法. 对于给定的优先搜索区域, 本文分别应用了标准差法, 集中化法, 以及概率距离法, 此外, 本文还设计了一种类似的算法, 这种算法认为近期的作案地点对于预测下一次作案可能的发生地点更为重要. 基于此, 本文提出了一种整合多种算法的结果的方法. 地理学分析方法的性能主要依赖于其是否能够有效的将一个特定的搜索区域缩小为一个更小的可能包含下一个作案地点的区域. 本文中所有方法产生的包含下一作案地点的优先搜索区域近似为整个搜索区域的 10%, 较均匀随机分布提高显著. 然而, 不同方法结果的差异在统计意义上并不显著. 同时, 我们还发现本文所有模型的精度还随着所研究的连环犯罪的不同而有所差异.

Contents

微信公众号：数学模型 MATHmodels

1 Introduction / 引言

Serial criminals present a unique challenge to law enforcement agencies. In a typical crime, investigators are able to draw a connection between the criminal and the victim. This information often provides enough clues to form the basis of a criminal investigation. In the case of serial criminals, however, there is usually no such relationship between the criminal and the victim [1, 2]. This lack of information about the serial criminal forces law enforcement agencies to consider a larger possible target area for the next crime, which hinders the investigation.

In order to better utilize limited law enforcement resources in a serial criminal investigation, geographical profiling can be used to determine likely next crime locations. Geographical profiling is “...a procedure that examines the spatial behavior of offenders with regard to the locations of their crime scenes and the spatial relationships between those scenes” [3]. By using geographical profiling, investigators are able to take advantage of the spatial patterns of a serial criminal to focus their attention on certain geographically important regions.

In this paper we examine different geographical profiling methodologies and compare their ability to predict the location of a serial criminal’s next crime. In Section 2, we provide the terminology and background information that will be utilized in the rest of the paper. In Section 3, we describe the features that are common to all geographical profiling methods developed and our metrics for determining the performance and accuracy of a particular method. Section 4 develops geographical profiling methods and reports on their performance, and Section 5 considers combinations of these methods. In Section 6, we apply our best geographical profiling methods to the serial murders committed by Peter Sutcliffe. Section 7 summarizes our main conclusions, while Section 8 discusses possible avenues for future work.

连环犯罪对于警方而言，是一个最为棘手的挑战. 对于普通的犯罪案件，调查者能够建立起作案者和受害者间的联系. 这些联系提供的线索通常足够形成调查犯罪的基础. 但是，对于连环犯罪而言，作案者和受害者间的这种联系通常并不存在 [1, 2]. 这种联系的缺失阻碍了调查，使得警方在确定下一次可能的作案地点时不得不考虑更大的目标区域.

在连环犯罪的调查中，为了更好的利用有限的警力，地理学分析可以被用来确定下一个可能的作案地点. 所谓地理学分析，就是指“ ...根据作案地点的空间位置以及作案地点间的空间关系，来推测出作案者在空间上犯罪行为的一种方法” [3]. 通过利用地理学分析技术，调查者能有效的利用连环犯罪在空间上的作案特点，从而将注意力集中在地理上某些重要的区域.

在本文中，我们研究不同地理学分析方法，并比较了它们在连环犯罪中预测下一次作案地点的能力. 在第2章节中，本文给出了相关术语和背景，这些术语和背景将在本文的后续部分被用到. 在第3章节中，本文描述了所有地理学分析方法的普遍特点，以及我们确定一个特定方法的性能和精度的测量方法. 在第4章节中，本文建立了一些地理学分析方法，并通过计算给出它们各自的性能，而在第5章节中，本文还考虑将多种算法结果整合起来 (作为新的结果). 在第6章节中，我们将本文中最好的地理学分析方法应用到 Peter Sutcliffe 的连环谋杀案中. 在第7章节中，本文总结出主要的结论，在第8章节中讨论了未来研究的方向.

微信搜一搜 数学模型

2 Problem Background / 问题背景

2.1 Terminology / 术语

- **Serial Criminal:** A serial criminal is a habitual offender who commits three or more related crimes over a span of time [1, 4]. We consider all forms of serial criminals together, such as murderers, rapists, arsonists, and burglars.
- **Spatiotemporal Data:** The locations and corresponding times of crimes committed by a serial criminal.
- **Spatial Behavior:** The spatial behavior of a serial criminal describes how the serial criminal chooses crime locations [5, 6]. We model the spatial behavior of a criminal based on the spatiotemporal data of their previous crimes.
- **Geographical Profiling Method:** A geographical profiling method is a particular methodology for modeling the spatial behavior of a serial criminal. We consider **spatial methods** that only consider the spatial data of previous crimes, **spatiotemporal methods** that also take into account the times of previous crimes, and **aggregate methods** that combine two or more geographical profiling methods.
- **Prioritized Search Area:** The area to search prioritized by the probability of the next crime given by a specific model. This area is important in determining how to allocate law enforcement resources.

- **连环犯罪:** 连环犯罪是指某一时间段内, 同一作案者 (惯犯) 连续犯有三起或三起以上的相关案件 [1, 4]. 本文综合考虑所有类型的连环犯罪, 例如, 杀人犯, 强奸犯, 纵火犯和盗窃犯.
- **时空数据:** 连环犯罪的各个作案地点及相应的作案时间.
- **空间行为 (特征):** 某一连环犯罪的空间行为描述了作案者是如何选择作案地点的 [5, 6]. 本文通过连环犯罪以往作案的时空数据来模拟犯罪分子的空间行为.
- **地理学分析方法:** 地理学分析方法是对连环犯罪的地理行为建模的一种特定的方法. 本文中的**空间方法**只考虑了以往作案的空间数据, 而**时空方法**同时还考虑了以住作案的时间, **整合化法** 则对两种或更多种地理学分析方法 (的结果) 进行了整合.

- **优先搜索区域:** 优先搜索区域由某一特定的模型给出空间上各个区域下次作案的概率后确定的, 优先搜索区域对于如何分配警力是非常重要的.

2.2 Survey of Current Literature / 文献调研

There is an ongoing debate as to the effectiveness of different geographical profiling methodologies. Rossmo uses spatial data to determine the residence of a criminal based on assumptions about spatial behavior, such as the tendency to commit crimes close, but not too close, to home [7]. As Rossmo's paper lacks a detailed description of his algorithm, and relies on empirically determined constants, it is difficult to reproduce his geographical profiling methodology accurately [8, 7, 3]. Van der Kemp and van Koppen survey

对于不同的地理学分析方法的效率, 还是一个饱受争议的问题. 根据“连环犯罪在选择作案地点时倾向于在离他们家比较近, 但是又不是非常近的地方的作案”这一假设, Rossmo 利用空间数据来确定犯罪分子的居住点 [7]. 由于 Rossmo 的论文中缺乏对他算法的详细描述, 同时依赖于经验参数的确定, 因此很难准确地重复出他的地理学分析方法 [8, 7, 3]. Van der Kemp 和 van

微信搜一搜 数学模型

and criticize theories of spatial behavior, including theories of crime location proximity to the criminal's residence and shortcomings of current geographical profiling methods [9]. Beauregard et al. and Snook et al. also examine the spatial behavior or serial criminals [10, 5]. Brown et al. take into account spatiotemporal data as well as additional features of the crime locations, such as proximity to highways, to predict crime locations [11, 2]. Recent work by O'Leary provides a mathematical foundation for geographical profiling [12].

We note that most of the work in the field of geographical profiling focuses on identifying the residence of a serial criminal. This allows law enforcement agencies to cross-reference the addresses of potential suspects to the predicted residency of the criminal in order to narrow their search. In contrast, this paper focuses on identifying probable next crime locations in a series of linked crimes. Being able to forecast crime locations would assist law enforcement agencies in ways predicting the residency of the criminal does not. For example, law enforcement agencies would be able to increase patrols along probable crime locations, or alert high-risk neighborhoods to the existence of the serial criminal. Since serial criminals often center their crime locations around their residence [10], these two problems are related, and thus we are able to adapt methods of identifying the residence to the problem of identifying next crime locations.

Koppen 在空间行为上的研究和批评理论，包括作案点接近居住点的理论以及对目前地理学分析方法的缺陷的批评. Beauregard 等人和 Snook 等人也对连环犯罪作了研究 [10, 5]. Brown 等人将更多的时空的数据 (比如靠近公路) 考虑进来，作为对作案地点特征的补充来预测下次作案地点 [11, 2]. O'Leary 在近期的研究成果则为地理学分析提供了一个数学基础.

我们注意到绝大多数地理学分析领域的方法都侧重于找出连环犯罪分子的居住点. 这种方法将允许警方参考潜在犯罪嫌疑人的住址来预测作案者的居住点位置，从而减小警方的搜索范围. 然而，本文却侧重于找出一系列相关的案件的下一次作案地点. 对作案地点地预测的能力能够协助警方来预测犯罪分子的居住地. 例如，警方可以在下次可能的作案地点区域附近增加巡逻的警力，或者在连环犯罪的高风险区附近发出警告. 由于连环犯罪的作案点都是以犯罪分子的居住点为中心 [10]，因此，这两个问题 (下一次作案地点和犯罪分子的居住点) 是相关的，因此我们可以通过修改预测犯罪分子居住点的模型来解决预测下一次作案地点的问题.

2.3 Collection of Data / 数据的采集

In order to determine the accuracy of proposed geographical profiling models, we needed to see how the models performed against actual spatiotemporal serial crime data. Unfortunately, there is no large collection of serial crime data to compare our models to, or a standard collection of serial crime data that is used across all geographical profiling research. A survey of spatial profiling research shows that the data sets used are varied, which leads to different quantitative conclusions about the spatial behavior of serial criminals; see [9, 8, 10, 5] for contrasting conclusions based on different data sets. For example, the circle center method described in Section 4.1 was found to have different levels of accuracy for serial crime data sets from different countries [4]. Ideally, we would be able to test our geographical

为了确定这些被提出的地理学分析模型的准确性，我们需要知道这些模型应用到实际的连环犯罪的时空数据时表现如何. 可惜的是，没有大量的连环犯罪的数据来测试我们的模型，也没有一个可用于所有地理学分析研究的普适的数据. 一份关于空间分析的调查显示用于地理学分析的数据是多种多样的，所使用数据的不同导致了不同的关于连环犯罪的空间行为的定量结论. 文献 [9, 8, 10, 5] 对使用不同数据得出的结论进行了对比. 例如，本文第4.1章节中所述的圆心法，对于来自不同国家的连环犯罪数据有着不同水平的准确程度 [4]. 理论上，我们可以用大量跨越多个地点，犯罪类型，时期的连

微信搜一搜 数学模型

profiling models against serial crime data that spanned multiple locations, crime types, and periods of time, but such a data setdoes not exist.

In most studies we considered, a sample of serial crime data was either obtained from a government agency or compiled from newspaper and police reports. Unfortunately, due to the constrained time frame of our study, we were unable to obtain data from police departments and government agencies. However, we were able to compile a data set consisting of the crimes of nine serial criminals, totaling 124 crimes. For the more notorious criminals in this data set including the Beltway Snipers, Peter Sutcliffe, and Dale Hausner we were able to collect data directly from police reports and news articles. We collected data for the remaining crimes from SpotCrime.com, an online crime information source, and verified the data via the referenced police reports and news articles [13]. A listing of the crimes in our data set can be found in Table 1.

环犯罪数据来测试我们的地理学分析模型, 但是这样的数据库并不存在.

在本文所考虑的大多的研究中, 连环犯罪的数据样本要么是从政府机构那获得, 要么是从报纸和警方的报告中获得的. 可惜的是, 由于研究时间的限制 (比赛时间有限), 我们不能从警方或者政府部门那获得数据. 不过, 我们仍然可以用 9 个连环犯罪, 共计 124 起案件组成一个数据库. 在这个数据库中, 包括 Beltway Snipers, Peter Sutcliffe, 以及 Dale Hausner 等臭名昭著的罪犯的相关数据, 是我们能够直接从警方的报告和新闻文章收集到的. 其余的数据是我们从 SpotCrime.com(一个犯罪信息的网络资源) 收集到的, 并通过警方的报告和新闻文章得到了核实 [13]. 本文关于连环犯罪的数据库的一个列表见表 1.

Description	Number of Crimes	Accuracy of Temporal Exponential Decay Model
Murders committed by Peter Sutcliffe in West Yorkshire, England	12	0.048
Beltway sniper attacks committed by John Allen Muhammad and Lee Boyd Malvo in the Washington, DC area	14	0.065
Murders, arsons, and other crimes committed by serial criminal Dale Hausner in Phoenix, AZ	40	0.098
Sexual assaults committed by a serial rapist in Columbus, OH	7	0.517
Serial robberies at TCF Banks in Frankin Park, IL area	12	0.180
Serial robberies committed in Denver, CO	6	0.103
Serial bank robberies committed by “The Withdrawal Bandit” in Boca Raton, FL	12	0.119
Serial robberies committed in Columbus, OH	7	0.020
Serial robberies and home invasions committed in Santa Monica, CA	14	0.174

Table 1: Description of data sets with accuracy score for temporal exponential decay model.

微信搜一搜 数学模型

3 Constructing a Model / 模型的建立

3.1 Assumptions / 假设

- **Spatiotemporal data is accurate:** A common modeling assumption.
- **Crimes were committed by one serial criminal:** Our models are working on the assumption that all the crimes in a data set were committed by the same criminal, and thus the crime locations can be explained by a particular spatial behavior.
- **Serial criminals do not act randomly:** We assume the truth of rational crime theory, the notion that “...there is an underlying reason why criminals choose to commit crimes at a particular time in a particular location” [2]. Rational crime theory is a necessary condition for productive geographical profiling, since if serial criminals are acting randomly, then the best model of their spatial behavior would be random as well.

- **时空数据是准确的:** 通常建模都要作的一个假设.
- **所有案件都是由同一个连环犯罪所犯:** 本文的模型都是基于“数据库中的所有案件都是由相同的连环犯罪所犯”这个假设来构建的. 因此可以用一个特定的空间行为来解释作案地点 (的选择).
- **连环犯罪并不会随机发生:** 我们假设理性犯罪理论是正确的, 这个理论是“...犯罪分子之所以选择在特定时间和特定地点的犯罪, 是有一个潜在的原因的” [2]. 理性犯罪理论是地理学分析成立的一个必要条件, 因为如果连环犯罪是随机发生的, 那么最好的关于连环犯罪的空间行模型也应当为随机的.

3.2 Model Descriptions / 模型的描述

Let $x_1, \ldots, x_n \in A \subset \mathbb{R}^2$ denote crime scene locations of a serial criminal, where A is a given search area encompassing the crime locations. Let $t_1, \ldots, t_n \in \mathbb{R}$ with $0 = t_1 < \ldots < t_n$ denote the corresponding times of the crimes, where (x_i, t_i) is the spatiotemporal data of the i^{th} crime. Given a location $y \in A$, we define a probability function P such that $P(y)$ is the probability that the next crime will happen at that location.

We discretize A into sectors $S_{i,j}$ in the order to simplify our calculations; see Figure 1 for a description of the discretization process. We assume that if $y \in S_{i,j}$, then $P(y) = P(S_{i,j})$, the probability that the next crime location is in the sector. This is a reasonable assumption for small sector sizes relative to the total search area. Thus, the goal of a geographical profiling method is to determine $P(S_{i,j})$ for all sectors.

For our models, let d be a given distance metric. This distance metric can be determined using a variety of methods, including standard Euclidean

令 $x_1, \ldots, x_n \in A \subset \mathbb{R}^2$ 表示一个连环犯罪的作案地点, 其中 A 表示一个给定的包含各作案地点的搜索区域. 令 $t_1, \ldots, t_n \in \mathbb{R}$ 且 $0 = t_1 < \ldots < t_n$ 表示相应的作案时间, 那么 (x_i, t_i) 就表示第 i 次作案的时空数据. 对于给定的空间位置 $y \in A$, 我们定义一个概率函数 P 使得 $P(y)$ 是下次作案地点发生在 y 处的概率.

为了简化计算, 本文将区域 A 离散成块 $S_{i,j}$; 离散化过程的描述见图 1. 本文假设如果 $y \in S_{i,j}$, 那么 $P(y) = P(S_{i,j})$, 其中 $P(S_{i,j})$ 为下次为作案地点落在该块中的概率. 这是一个合理的假设, 因为每一小块相对于整个搜索区域来讲很小. 因此, 地理学分析方法的目标就是求所有块中的 $P(S_{i,j})$.

令 d 为一种给定的距离度量. 这种距离的度量可以由多种方法确定, 其中包括标准的欧氏距离, 曼哈顿距

微信搜一搜 数学模型

distances, Manhattan distances, or travel-time distances that take into account the road map data of the area. We arbitrarily choose to use Euclidean distances since we found no evidence that serial criminals evaluate distance using a certain metric, but our models do not make any assumptions about the nature of the distance function.

离，旅行时间距离 (要考虑某地区的道路地图数据). 由于我们并没有发现有任何证据表示连环犯罪的距离是由哪一种特定的方法确定的，不失一般性，这里我们选用了欧氏距离，但本文的模型对距离函数的性质并没有作任何假设 (意思是本文的模型对其它距离也适用).

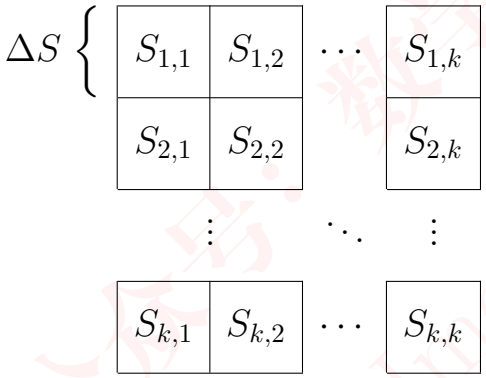


Figure 1: The discretization of a square search area into sectors. In this figure k is equal to the side length of the search area divided by Δs.

3.3 Model Performance Metric / 模型性能的评价指标

There exist two commonly used metrics for the evaluation of the performance of a geographical profiling method. The first measure is the error distance, which is the Euclidean distance between the most likely exact location of the next crime and the actual next crime location [8]. The second measure represents how much of the prioritized search area would need to be searched in order to find the next crime location [12, 7].

Our model evaluation metric should reflect how useful these models would be to a law enforcement agency tracking a serial criminal. The most likely exact location of the next crime is not very valuable to law enforcement agencies, since often they are concerned with finding an area to search and not a particular point [12, 14, 7].

Thus, we consider the following metric that represents how much of

评价地理学分析法性能常用的衡量指标有两种. 第一种衡量指标为误差距离，也就是最可能的下次作案位置与实际作案位置之间的欧氏距离 [8]. 第二种为衡量指标则表示：为了搜索到下次作案的实际位置，优先搜索的区域需要有多大 [12, 7].

我们对模型的评价指标必需反应这些方法对警方追踪连环犯罪有多大作用. 下次最有可能作案的位置对警方而言并没有太大的意义，因为通常他们更关心 (如何) 确定一个区域来搜索，而不是一个确定的位置点 [12, 14, 7].

因此，本文考虑后一种衡量指标：为了搜索到下次作

微信搜一搜 数学模型

the prioritized search area would need to be searched in order to find the next crime location, which we call the **hit score**. Given a geographical profiling method with associated probability function P and a known next crime location x_{n+1}, we wish to determine the hit score H.

Let S denote the set of all sectors. Let $L \ni x_{n+1}$ denote the sector containing the next crime location. Then

$$B = \{S_{i,j} \in S : P(S_{i,j}) > P(L)\}$$

is the set of all sectors that have a higher predicted probability than the sector containing the next crime. Thus, we have

$$H = \frac{|B|}{|S|}$$

which is the fraction of the search area that would need to be searched in prioritized order before finding the next crime location. Note that a lower value for the hit score is preferable.

3.4 Model Accuracy for a Particular Criminal / 模型的准确度

Given an active serial criminal, we wish to determine the accuracy to which the criminal's spatial behavior is determined by a model. This is important information for law enforcement agencies basing investigative decisions on a geographical profiling method, since they want to know the extent to which they can trust a model as it is applied to a particular criminal.

We calculate model accuracy for a given serial criminal in the following way. Given a geographical profiling method with associated probability function P, we wish to determine the **accuracy score** of the method, denoted Z.

Let H_k denote the hit score that considers $x_1, \ldots, x_{k-1}$ as the currently known crime locations, and treats x_k as the next crime location. Then,

$$Z = \frac{1}{n-3} \sum_{k=4}^{n} H_k$$

is the mean of the hit scores determined by sequentially adding each crime

案的实际位置, 优先搜索的区域需要有多大, 本文中称为**命中比分**. 对于给定的地理学分析法及相应的概率函数 P 和已知的下次作案位置 x_{n+1}, 本文希望能够确定出命中比分 H.

令 S 表示所有块的集合. 令 $L \ni x_{n+1}$ 表示包含下次作案位置的块. 那么

$$B = \{S_{i,j} \in S : P(S_{i,j}) > P(L)\}$$

就是概率比下次作案点实际所在块高的所有块的集合. 因此, 我们可以用函数

$$H = \frac{|B|}{|S|}$$

来表示在找到下次作案位置前, 按照优先顺序需要被搜索的区域所占比例. 值得注意的是: 命中比分越小越好.

对于给定的连环犯罪, 本文期望确定由某种方法算得的空间行为的准确性. 这对于凭某种地理学分析方法做出决策的警方是非常重要的信息, 因为他们想知道将某一模型应用到一个特定的案件时, 在何种程度上可以信任模型 (的结果).

我们将用以下的方式来计算在某一给定连环犯罪下模型的准确度. 对于一个给定的地理学分析方法及相应的概率函数 P, 本文期望确定这个方法的**准确度得分** Z.

假设目前已知的连环犯罪作案位置 $x_1, \ldots, x_{k-1}$, 并将 x_k 作为下个作案点位置, 令 H_k 表示命中比分. 那么,

$$Z = \frac{1}{n-3} \sum_{k=4}^{n} H_k$$

表示从第四个作案点开始, 依次添加每个作案地点, 直

微信搜一搜 数学模型

location, starting with the fourth, to the set of currently known crimes. We start with the fourth crime because criminal behavior is not generally considered serial until the criminal has committed at least three crimes [1]. Note that a lower value for the accuracy score represents a more accurate model.

到当前已知的最后一个作案点为止的命中比分平均值. 我们从第四次作案点算起是因为犯罪分子的空间行为只有在三次以上的作案中一般才会被考虑 [1]. 需要注意的是: 准确度得分越低, 表示模型的准确度越高.

3.5 Search Area Determination / 搜索范围的确定

Our notions of performance and accuracy depend heavily on the size of the search area. The hit score can be made arbitrarily small by increasing the size of the search area. This is because additional locations on the periphery of the search area are unlikely to be the next crime location, but these additional locations are included in the total search area.

Surprisingly, literature that uses the hit score performance metric does not explicitly address how the search area is to be determined given previous crime locations. We take the search area to be a square centered at the mean location of the previous crimes, with side length equal to double the maximum pairwise distance between previous crimes. In an ongoing serial criminal investigation the search area could be provided by law enforcement, but for our purposes the above search area balances the size of the search area based on previous crime distances.

本文中的性能和准确度的概念强烈依赖于搜索范围的尺寸. 通过增加搜索范围, 可以使命中比分变得任意的小. 这是因为在搜索区域的边缘上额外增加的区域都不太可能成为下一个作案地点, 但这些额外增加的区域却要被包括在总搜索区域中.

奇怪的是, 用到命中比分这个性能指标的文献中却没有明确阐述如何确定在给定以往作案点下的搜索范围. 本文将以往作案位置的平均值为中心, 并以两倍于最远的两个作案位置的距离为边长的方形区域作为搜索范围. 对于一个正在被调查的连环犯罪, 搜索的范围可以由警方提供, 但是对本文而言, 上述的搜索区域与基于以往作案点之间距离的搜索区域尺寸是相对应的.

4 Individual Models / 单一的模型

4.1 Spatial Models / 空间数据模型

In this section we consider geographical profiling methods that only take into account the spatial data of serial crimes.

We consider the following spatial models for the calculation of $P(S_{i,j})$:

- **Random Method:** The random method assigns each $P(S_{i,j})$ a random value uniformly. Theoretically, we expect the hit score of the random method to be 0.5 [14]. We include the random method as a basis for comparison of other geographical profiling methods.
- **Standard Deviation Methods:** Since standard deviation methods give no information about the distribution of probabilities inside the

在这个子章节中, 我们所考虑的地理学分析方法都只涉及连环犯罪的空间数据.

我们考虑下面的空间模型来计算 $P(S_{i,j})$:

- **随机法:** 随机法随机的对每个 $P(S_{i,j})$ 赋一个均匀分布的随机值. 理论上, 我们可以得到随机法的命中比分的期望值为 0.5 [14]. 本文将随机法作为与其它地理学分析方法比较的基础.
- **标准差法:** 由于标准差法并不能给出标准差范围内概率分布的有关信息, 因此将它们同其它方法比

微信搜一搜 数学模型

standard deviation areas, we cannot meaningfully compare them to other methods. Regardless, we include standard deviation methods in this paper because they are the most basic geographical profiling methods. These methods provide a rudimentary way by which the potential search area can be narrowed. We consider the following standard deviation methods:

- **Standard Deviation Rectangles:** These are rectangular areas defined by the points

$$\begin{aligned} \overline{x} &+ (-c\,\sigma_{lon}, -c\,\sigma_{lat}) \\ \overline{x} &+ (-c\,\sigma_{lon}, c\,\sigma_{lat}) \\ \overline{x} &+ (c\,\sigma_{lon}, c\,\sigma_{lat}) \\ \overline{x} &+ (c\,\sigma_{lon}, -c\,\sigma_{lat}) \end{aligned}$$

  where $\overline{x}$ is the centroid of the crime locations, and σ_{lon} and σ_{lat} are the standard deviations of the longitudes and latitudes of the crime locations [15].
- **Standard Deviation Ellipses:** These are elliptical areas, oriented along the trend line of the data in the least-squares sense. A standard deviation ellipse is an ellipse with its center at the centroid of the crime locations, rotated clockwise by an angle θ, and with axis lengths given by $2c\sigma_{lon}$ and $2c\sigma_{lat}$, where θ, σ_{lon}, and σ_{lat} are calculated as in [16].

In both of the above methods, c is a constant that determines the range of the area. Common values are $c = 1$ for the 68^{th} percentile area and $c = 2$ for the 95^{th} percentile area. See Figure 7 for an illustration of these percentile areas.

- **Centralization Methods:** Centralization methods determine a central focal point for the spatial pattern of the serial criminal. In these models, the probability of the next crime decreases the further the location is from the focal point. Thus, in all these methods, given a central focal point $C \in A$, we have

$$P(S_{i,j}) = \frac{1}{d(S_{i,j}, C)}$$

较是没有意义的. 但是我们仍然将标准差法包括在了本文中, 这是因为该算法是最基本的地理学分析法. 这些方法提供了一个缩小潜在的搜索区域的方法. 本文考虑以下标准差法:

- **标准差矩形:** 这些矩阵都是通过顶点定义的

$$\begin{aligned} \overline{x} &+ (-c\,\sigma_{lon}, -c\,\sigma_{lat}) \\ \overline{x} &+ (-c\,\sigma_{lon}, c\,\sigma_{lat}) \\ \overline{x} &+ (c\,\sigma_{lon}, c\,\sigma_{lat}) \\ \overline{x} &+ (c\,\sigma_{lon}, -c\,\sigma_{lat}) \end{aligned}$$

  其中 $\overline{x}$ 是作案地点的中心, σ_{lon} 和 σ_{lat} 是作案地点经纬度的标准差 [15].
- **标准差椭圆:** 沿着数据满足最小二乘的方向的椭圆. 一个标准差椭圆是一个以作案地点的质心为中心, 顺时针旋转一个角度 θ, 长短半轴分别为 $2c\sigma_{lon}$ 和 $2c\sigma_{lat}$ 的椭圆. 其中 θ, σ_{lon} 和 σ_{lat} 的计算方法见 [16].

在以上两种方法中, c 都是决定区域范围的一个常数. 对于 68% 的面积, 取值为 $c = 1$, 对于 95% 的面积, 取值为 $c = 2$ 图7是这些百分面积的一个示意图.

- **中心化方法:** 中心化方法给空间上分布的连环犯罪的作案点确定了一个中心焦点. 在这些模型中, 下次作案地点的概率从中心焦点往外减小. 因此, 在所有的这些方法中, 给定的中心焦点 $C \in A$, 本文有

$$P(S_{i,j}) = \frac{1}{d(S_{i,j}, C)}$$

微信搜一搜 数学模型

We consider the following ways of determining the central focal point:

- **Centroid:** The central focal point is the mean of the locations of the crimes, given by
$$C = \frac{1}{n}\sum_{i=1}^{n} x_i$$
- **Harmonic Mean:** The central focal point is the harmonic mean of the locations of the crimes, given by
$$C = \frac{n}{\sum_{i=1}^{n}\frac{1}{x_i}}$$
- **Circle Center:** The central focal point is the mean location of the two crime locations farthest away from each other. This is determined as follows. Let x_i, x_j be such that $d(x_i, x_j)$ is maximal. Then we have $C = \frac{x_i + x_j}{2}$.
- **Median:** The central focal point is defined as the point that is the median of the of the longitudes and the median of the latitudes of the crime locations. Compared to the other centralization methods, the median is less sensitive to distant crime locations, which the criminal might have chosen for no particular spatial reason [9].

- **Probability Distance Method:** The probability distance method takes into account a location's distance from each particular previous crime location. In this model, a probable next crime location would be relatively close to multiple previous crimes. Thus, we have
$$P(S_{i,j}) = \sum_{k=1}^{n} f(d(S_{i,j}, x_k))$$
where f is a distance decay function defined in one of the following ways:
  - **Linear Distance Decay:** The probability of the next crime location decreases linearly away from a particular previous crime location, given by $f(d) = \alpha - \beta d$, where α, β are decay constants.

本文考虑以下方法来确定中心焦点:

- **形心:** 用作案位置的平均值作为中心焦点, 即
$$C = \frac{1}{n}\sum_{i=1}^{n} x_i$$
- **调和平均值:** 用作案位置的调和平均值作为中心焦点, 即
$$C = \frac{n}{\sum_{i=1}^{n}\frac{1}{x_i}}$$
- **圆心:** 用两个最大距离的作案位置的平均值 (中点) 作为中心焦点. 具体是确定办法如下: 令 x_i, x_j 为两个作案位置并有 $d(x_i, x_j)$ 最大, 那么有 $C = \frac{x_i + x_j}{2}$.
- **中位数:** 用作案位置的经纬度的中位数构成的点作为中心焦点. 与其它中心化方法相比, 中位数对犯罪分子可能选择的远处作案点的位置不太敏感 (犯罪分子可能在远处作案, 而没有什么特殊的空间原因 [9]).

- **概率距离法:** 概率距离法将一点到各个以往作案地点的距离考虑了进来. 在这个方法中, 下个可能的作案地点的位置会相对比较靠近较多的以往作案地点. 因此, 本文有
$$P(S_{i,j}) = \sum_{k=1}^{n} f(d(S_{i,j}, x_k))$$
其中 f 是距离衰减函数, 由下列方式之一定义
  - **线性距离衰减:** 某个位置是下个作案位置的概率随着与以往作案位置的距离线性衰减, 衰减形式为 $f(d) = \alpha - \beta d$, 其中 α, β 为常数.

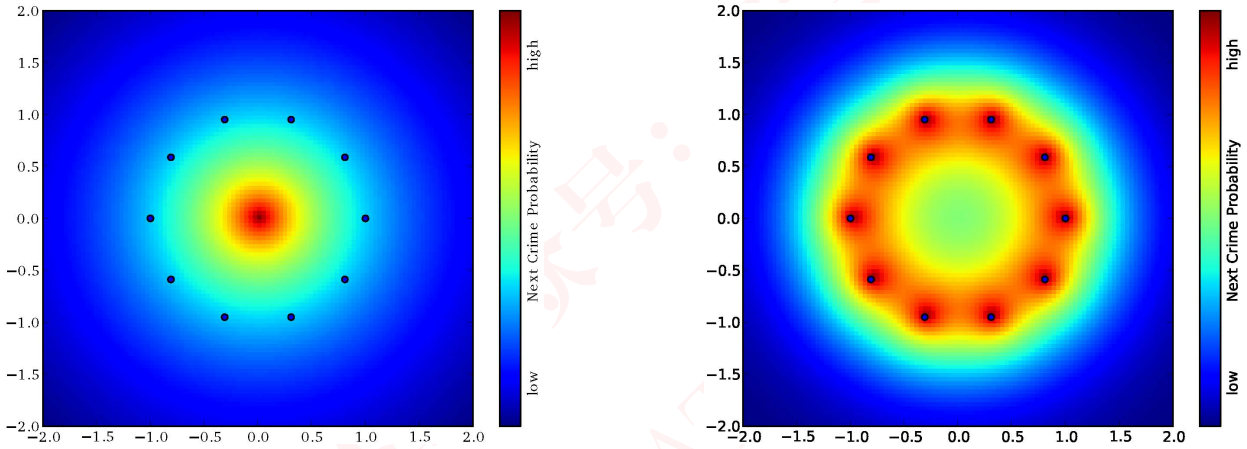
微信搜一搜 数学模型

– **Exponential Distance Decay:** The probability of the next crime location decreases exponentially away from a particular previous crime location, given by $f(d) = e^{-\gamma d}$, where γ is a decay constant.

For an illustrative example of the difference between centralization methods and the probability distance method see Figure 2.

– **指数距离衰减:** 某个位置是下个作案位置的概率随着与以往作案位置的距离指数衰减, 衰减的形式为 $f(d) = e^{-\gamma d}$, 其中 γ 为常数.

一个中心化方法和概率距离法之间差异的例子见图 2.



(a) Centroid centralization method on the unit circle

(b) Exponential decay method on the unit circle

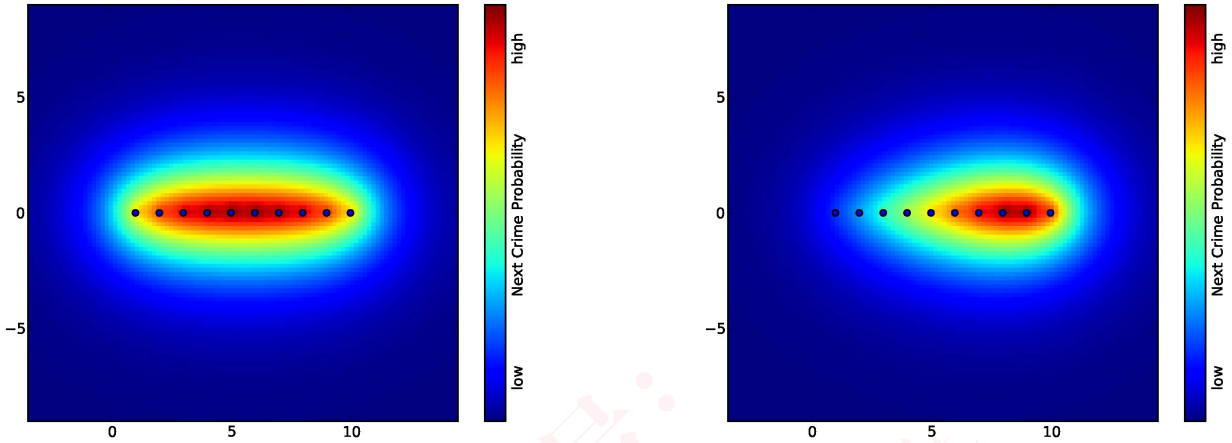
Figure 2: Comparison of prioritized search areas of a hypothetical criminal committing crimes along the unit circle given by a centralization method and a decay method.

4.2 Spatiotemporal Models / 时空模型

In this section we consider geographical profiling methods that take into account both the spatial and temporal data of serial crimes. The additional use of temporal data is motivated by the idea that recent crime locations are more relevant to the spatial behavior of a serial criminal than older crime locations. For example, if the serial criminal is traveling while committing crimes, then old crime location information quickly becomes outdated. For

在本章节中, 本文研究的地理学分析方法将同时考虑连环犯罪的空间数据和时间数据. 时间数据的额外使用是出于“时间上最近的作案地点比其它作案地点更加与连环犯罪的空间行为相关”这样的一个想法. 例如, 如果连环犯罪的作案者一边作案一边迁移, 那么旧的作案点信息将很快过时. 关于这种情况的一个例子, 见图 3.

微信搜一搜 数学模型



(a) Exponential Decay (b) Temporal Exponential Decay

Figure 3: Prioritized search area for a hypothetical serial criminal travelling east over time, as given by the exponential decay and temporal exponential decay models.

an example of this, see Figure 3.

Not all of the spatial models in Section 4.1 have extensions that incorporate temporal data. For example, the circle center centralization method has no spatiotemporal analog, since the longest pairwise distance is unaffected by temporal data.

To incorporate the temporal components of the crime data, we calculate a **temporal weighting factor** for each crime,

$$w_i = \frac{t_i - t_1}{t_n} + k$$

where w_i denotes the temporal weight of the i^{th} crime. The offset k is included so that the first crime will not be given a weight of 0. We chose $k = 0.1$ so that the last crime is weighted approximately ten times more heavily than the first crime. Let W denote the sum of the temporal weights.

We modify the following spatial models to consider the temporal weighting factors:

4.1章节的所有空间模型并非都能通过扩展, 来引入时间数据. 比如, 由于作案位置两点间的最大距离并不受时间数据的影响, 圆心中心化方法中并不涉及时空关系.

为了加入犯罪数据中的时间组份, 本文计算了每一次作案的**时间权重因子**.

$$w_i = \frac{t_i - t_1}{t_n} + k$$

其中 w_i 表示第 i 次作案的时间权重. k 是补尝量, 被加入进来是为了使第一次作案的权重不会被赋值为 0. 本文取 $k = 0.1$, 使得最后一次作案的权重比首次作案的约高十倍左右. W 表示时间权重的总和.

我们通过改进以下的空间数据模型来考虑时间加权因子:

微信搜一搜 数学模型

- **Centralization Methods:**
  - **Temporal Centroid:** The mean of the locations of the crimes, weighted by time, given by

$$C = \frac{1}{W}\sum_{i=1}^{n} x_i w_i$$

  - **Temporal Harmonic Mean:** The harmonic mean of the locations of the crimes, weighted by time, given by

$$C = \frac{W}{\sum_{i=1}^{n} \frac{w_i}{x_i}}$$

- **Temporal Probability Distance Methods:** The linear distance decay and exponential distance decay methods incorporate the temporal weight data by using the following modified probability function:

$$P(S_{i,j}) = \sum_{k=1}^{n} w_k f(d(S_{i,j}, x_k))$$

- **中心化方法:**
  - **时间形心法:** 通过加权时间的作案地点的平均值定义如下

$$C = \frac{1}{W}\sum_{i=1}^{n} x_i w_i$$

  - **时间调和平均:** 通过加权时间的作案地点的调和平均值定义如下

$$C = \frac{W}{\sum_{i=1}^{n} \frac{w_i}{x_i}}$$

- **时间加权概率距离法:** 考虑时间加权数据的线性距离衰减和指数距离衰减是通过以下修正的概率方程实现的:

$$P(S_{i,j}) = \sum_{k=1}^{n} w_k f(d(S_{i,j}, x_k))$$

4.3 Individual Model Results / 单一模型的结果

To investigate the effectiveness of each of the aforementioned models, we calculated the mean hit score across all of our data sets. After three crimes have been observed we apply our model and record the hit score for the next crime. This process is repeated for all subsequent crimes in the series.

The value of the constants in the linear and exponential decay models do not effect the hit score. This is because the hit score is not affected by the magnitude of the probability of the sector but by the relative probability to the rest of the search area. Consequently for an individual model the values of the constants have no effect.

First we investigate the overall performance of each of the models. This can be found in Table 2 and in Figure 4. All of our models significantly outperform the random model, but each of the individual models is not sta-

为了研究上述模型的效力, 我们计算了我们的数据库中所有案例的平均命中比分. 以前三次作案作为已发生的, 我们应用本文的模型逐个记录下一次作案的命中比分. 对于一个连环犯罪随后的所有子犯罪 (从第四次子犯罪开始到最后的子犯罪), 不断重复此过程.

线性和指数衰减模型中的常数取值不会影响命中比分. 这是因为命中比分并不受每块概率的绝对大小的影响, 而受与其它搜索区域概率的相对比值影响. 因此对于一个单独的模型, 常量的取值没有影响.

首先, 我们研究了所有模型的整体性能. 这些性能指标见表2 和图 4. 本文中的所有模型都明显优于随机模型, 但在统计上任一单独的模型并没有绝对的优于其它

微信搜一搜 数学模型

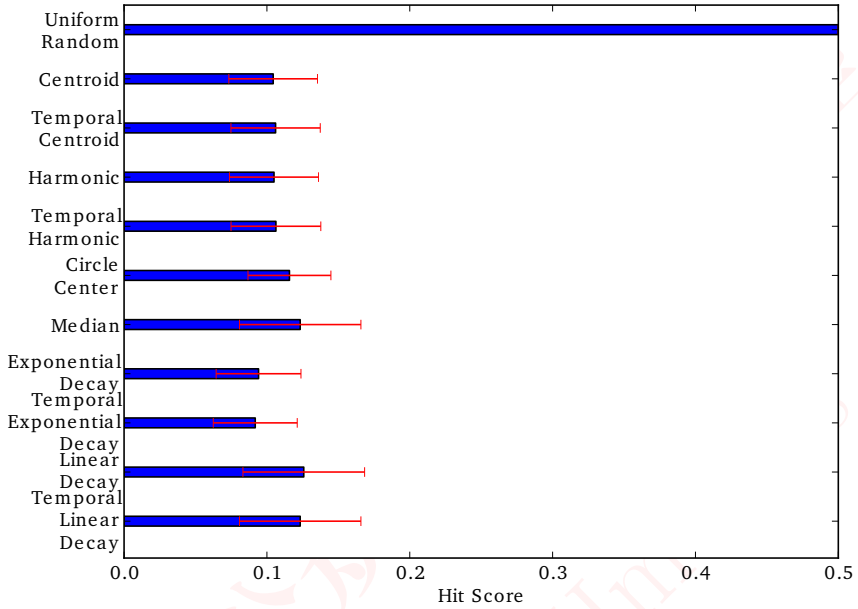


Figure 4: Comparison of the relative performance of each of our models on all of our data sets.

tistically superior to any other.

Next we investigate the performance of our models based upon the number of crimes observed as seen in Figure 5. No patterns emerge that would indicate that the performance of the model improves having observed more crimes. The data sets considered do not have sufficient data to rule out such a correlation, particularly with the longer streaks as there are only a few data points.

(所有模型的结果差不多, 没有特别出众的).

接着我们根据不同已知的作案数量研究了本文模型的性能, 具体见图 5. 并没有出现任何迹象表明当已知的作案数更多时模型表现更佳. 数据库没有足够的数据以排除这种关联, 特别是较长的质信区间内只有少量的几个数据点.

5 Aggregation Models / 整合模型

In this section we consider ways of combining predictions from multiple geographical profiling methods into one model. The motivation behind these aggregate approaches is that each of the individual models described in Sec-

在这一章节中, 我们考虑一种将多种地理学分析法预测的结果整合成一个模型的方法. 采用整合方法的潜在动机是, 在第4.1和4.2章节中的任何一个单一模型都有

微信搜一搜 数学模型

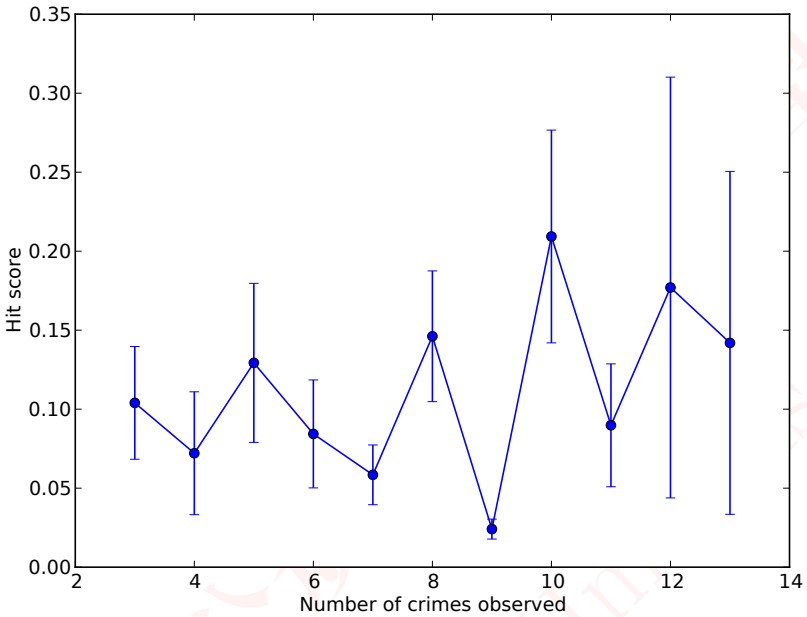


Figure 5: Hit score by number of crimes observed using a exponential decay model and 50% confidence intervals.

tions 4.1 and 4.2 have strengths and weaknesses, but by aggregating several models we hope enhance the strengths and reduce the weaknesses.

This issue of aggregation of models relates to the classic problem of aggregating expert predictions. Consequently, there exist several different techniques for aggregating these models. These techniques can be categorized as either axiomatic or Bayesian. Clemen and Winkler reported no significant difference in the performance of more complex Bayesian models with more simple axiomatic approaches [17], and as such we have chosen to focus on the more elementary axiomatic approaches.

Axiomatic Approaches: Given n models that we wish to aggregate, let $P_k(S_{i,j})$ denote the probability that a particular sector contains the next crime as predicted by the k^{th} model. Let W_k denote the aggregation weight of the k^{th} model, given such that all W_k's sum to one.

优点和缺点, 但是通过整合多种模型, 我们希望增强优势而减小劣势.

这种组合模型的方式涉及到一个经典问题: 如何整合多位专家的预测. 因此, 存在多种不同的整合模型的方法 (整合专家预测的方法已有多种). 这些方法可以被分类为自适应方法和贝叶斯方法. Clemen 和 Winkler 指出特别复杂的贝叶斯方法与简单的自适应方法没有太明显的差异 [17], 因此我们选择把重点放在了更基本的自适应方法上.

自适应方法: 给出我们希望整合的 n 种模型, 令 $P_k(S_{i,j})$ 表示第 k 种模型预测 $S_{i,j}$ 块包涵下次作案地点的概率. 令 W_k 表示每 k 种模型的整合权重, 并且所有的 W_k 的总和为 1.

微信搜一搜 数学模型

Model	Hit Score	95% Confidence
Random	0.500	0.001
Centroid Centralization	0.104	0.031
Temporal Centroid Centralization	0.106	0.031
Harmonic Mean Centralization	0.105	0.031
Temporal Harmonic Mean Centralization	0.106	0.031
Circle Center Centralization	0.116	0.029
Median Centralization	0.123	0.042
Exponential Decay	0.094	0.030
Temporal Exponential Decay	0.092	0.029
Linear Decay	0.125	0.043
Temporal Linear Decay	0.123	0.043

Table 2: Models with associated hit scores and 95% confidence intervals.

- **Linear Opinion Pool:** This model is a linear combination of two or more probability distribution functions produced by individual models.

$$P(S_{i,j}) = \sum_{k=1}^{n} W_k P_k(S_{i,j})$$

Examples of this model and the effect of different weights can be found in Figure 6.

- **Logarithmic Opinion Pool:** This model is a weighted product of two or more probability distribution functions.

$$P(S_{i,j}) = \prod_{k=1}^{n} P_k(S_{i,j})^{W_k}$$

In the exponential decay and the linear decay spatial models described in Section 4.1, we noted the value of the constants chosen for the model has no impact on the hit score metric. In an aggregate model this is not the case, as the relative magnitude of probability is important when added or multiplied by another model.

- **线性组合:** 这个方法将两个或两个以上的不同模型的概率分布函数线性的组合在一起.

$$P(S_{i,j}) = \sum_{k=1}^{n} W_k P_k(S_{i,j})$$

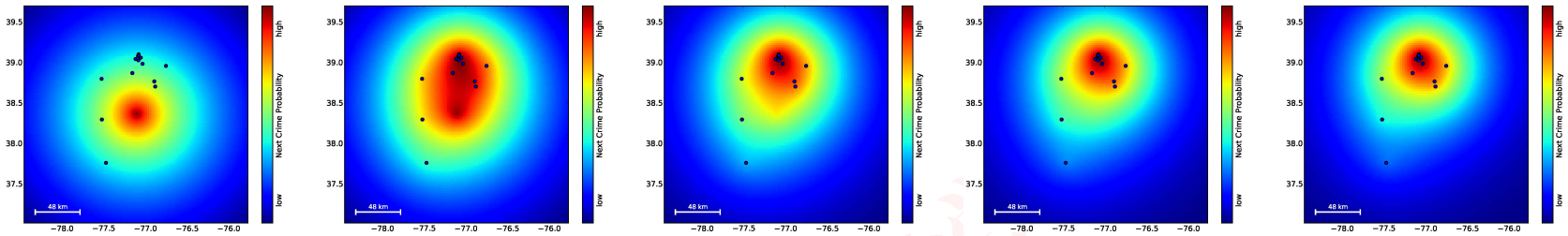
关于这个模型以及不同权重的效果的一些例子见图6.

- **对数组合:** 这个方法是将两个或两个以上的不同模型的概率分布函数进行加权乘积

$$P(S_{i,j}) = \prod_{k=1}^{n} P_k(S_{i,j})^{W_k}$$

在第4.1章节中的指数衰减和线性衰减模型中, 我们意识到数常值的选取对模型的命中比分没有影响. 但在整合模型中并不是这样, 这是因为相对概率的大小在与另一个模型的相加或相乘时是非常重要的.

微信搜一搜 数学模型



(a) Exponential decay weight 0.00, Circle center weight 1.00

(b) Exponential decay weight 0.25, Circle center weight 0.75

(c) Exponential decay weight 0.50, Circle center weight 0.50

(d) Exponential decay weight 0.75, Circle center weight 0.25

(e) Exponential decay weight 1.00, Circle center weight 0.00

Figure 6: Prioritized search areas resulting from using different weights in the aggregation model of a circle center and exponential decay for the Beltway Sniper.

To set these parameters we used the mean pairwise distance. We define the mean pairwise distance to be the mean distance between any two crimes

$$\overline{\delta} = \frac{2}{n(n-1)} \sum_{1 \le i < j \le n} d(x_i, x_j)$$

In the case of the linear decay model, α was chosen to be 1 and β was chosen such that $\alpha - \beta\overline{\delta} = 0$. In the case of the exponential decay, we set $\gamma = \sqrt{\frac{2}{\overline{\delta}}}$, which gives a mean distance to crime of $\frac{\overline{\delta}}{2}$. We believed these values would provide a reasonable magnitude of probability when compared with other models.

为了设置用来计算两点之间平均距离的参数. 我们定义平均两作案点之间的距离为任何两作案点之间距离的平均值.

$$\overline{\delta} = \frac{2}{n(n-1)} \sum_{1 \le i < j \le n} d(x_i, x_j)$$

在性线衰减的情况下, α 的值被选取为 1, β 的值的选取满足 $\alpha - \beta\overline{\delta} = 0$. 在指数衰减的情况下, 本文令 $\gamma = \sqrt{\frac{2}{\overline{\delta}}}$, 这将使得平均作案距离为 $\frac{\overline{\delta}}{2}$. 我们认为当与其它模型相比较时, 这些取值将提供一个合适大小的概率.

5.1 Aggregation Model Results / 整合模型的结果

To investigate the benefits of aggregation models, we used a 2^k **factorial** experimental design [18]. For each of our models described in Section 4, we investigated two possible scenarios: the model was included in the aggregate and weighted equally with the others, and the model was not included. We

为了研究整合模型的优势, 我们使用了一个2^k 析因试验设计[1] [18]. 对于第4章节中我们描述的每一个模型, 我们研究了两种可能的情况: 一种包括该模型的等权重加权, 另一种是不包含该模型 (的等权重加权). 然后我

[1]析因试验设计: 是一种将两个或多个因素的各水平交叉分组, 进行实验 (或试验) 的设计. 它不仅可以检验各因素内部不同水平间有无差异, 还可检验两个或多个因素间是否存在交互作用. 若因素间存在交互作用, 表示各因素不是独立的.

微信搜一搜
数学模型

then computed the mean hit score for the resultant aggregate model for all possible models on all of our data sets.

While a further optimization process could be conducted in which more weights or different decay constants were investigated, this process allowed us to see the effect of different models on the aggregate as well as see which individual models produce a strong aggregate model.

Through this process we found the strongest combination of individual models used the temporal exponential decay and the circle center centralization each equally weighted. The mean hit score for the linear opinion pool was 0.081 ± 0.027 with a 95% confidence interval. The logarithmic opinion pool did not perform any better than any individual model. Neither model is statistically better than any of our other models.

我们对本文的整个数据库中所有可用的模型产生的所有整合模型, 计算平均命中比分.

而进一步的优化过程中, 可以进行不同权重以及不同衰减常数的研究, 这个过程可能使我们看到不同模型在整合中的影响, 以及哪些模型可以生成强大的整合模型.

通过这个过程, 本文发现最强的模型的组合是等权重的时间指数衰减和圆心中心化方法. 这种线性组合的平均命中比分在 95% 的质信区间时为 0.081 ± 0.027. 对数组合并没有比其它模型表现出任何更优越的地方. 这两种模型在统计上也没有比其它模型更好.

6 Peter Sutcliffe: A Case Study / Peter Sutcliffe: 一个算例

Peter Sutcliffe was a serial murderer who targeted women in England during the late 1970's [19]. A map of his murders can be found in Figure 8(a). We have applied the models we have developed to predict where the next murder in the series would be if Sutcliffe had not been arrested and imprisoned.

A common but naïve method of identifying an area of high probability of serial crime generates a rectangle or ellipse in which a standard deviation of previous crimes have taken place. These rudimentary models can be seen in Figure 7.

In Section 4.3, we found the individual model with the lowest mean hit score across our data set was the temporal exponential decay model. See Figure 8 for an application of this model to the data for murders committed by Peter Sutcliffe. The known crimes had a hit score of 10.6% using this model, so we use this percentage to find a prioritized search area where we can be reasonably confident Sutcliffe would attack next. Of the $27818\,\mathrm{km}^2$, we can then isolate $2941\,\mathrm{km}^2$ in which to concentrate the search effort.

Peter Sutcliffe 是二十世纪 70 年代末英国的一起连环杀人案的凶手, 他的作案目标为妇女 [19]. 一张他作案点的地图如图 8(a)所示. 本文应用我们建立的模型来预测在这起连环犯罪中下一次作案地点将会在哪, 如果 Sutcliffe 还没有被逮捕并监禁.

一种简单而常用的鉴别连环犯罪高概率发生区域的方法是根据以发生案件的作案点的标准差产生一个矩形或椭圆. 这两个初步模型的示意图见 7.

在第 4.3 章节中, 对于本文数据库, 我们发现平均命中比分最低的单一模型是时间指数衰减模型. 图 8 是将这一模型应用到 Peter Sutcliffe 连环杀人案中的结果. 利用这个模型可以得到这个著名连环犯罪的命中比分为 10.6%, 因此我们用这个百度比来寻找一个我们有理由相信将发生下次作案的优先搜索区域. 对于 $27818\,\mathrm{km}^2$ 的区域, 我们可以隔离出 $2941\,\mathrm{km}^2$ 的面积以集中精力搜索.

微信搜一搜 数学模型

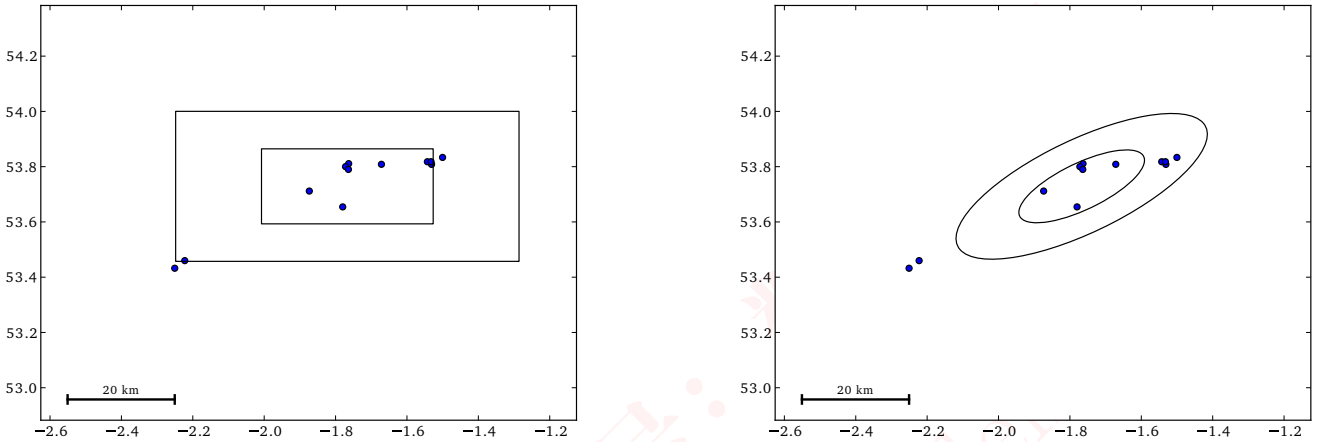


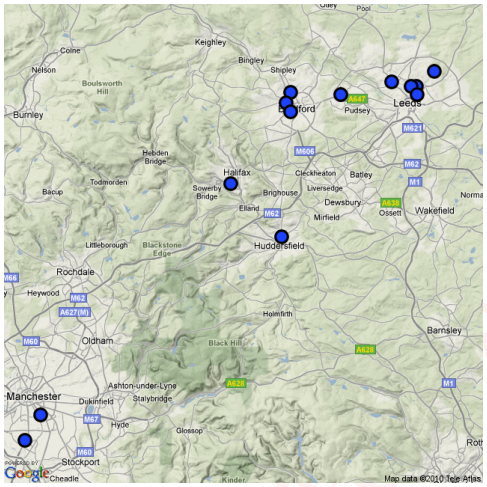
Figure 7: Standard deviation rectangles and ellipses for the murders committed by Peter Sutcliffe.

7 Conclusions / 结论

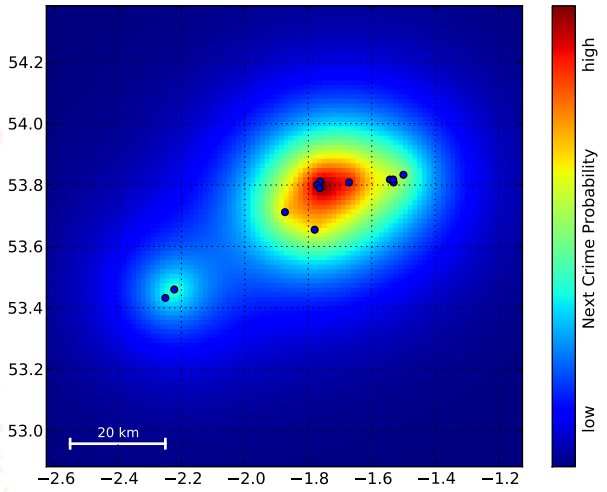
- **Every geographical profiling method outperformed the random method:** On our data set, every geographical profiling method provided a significant improvement in hit score over the random method.
- **All non-random geographical profiling methods considered exhibited roughly the same performance:** Differences in the hit score for different spatial, spatiotemporal, and aggregation methods were statistically insignificant, indicating that all the geographical profiling methods have the same level of performance. This is a similar conclusion reached by Snook et al., who show that complex geographical profiling methods are no more accurate than simple geographical profiling methods [4].
- **Particular geographical profiling methods are not applicable to all serial criminals:** We can see in Table 1 that the accuracy of our best individual geographical profiling method varied largely depending on the particular serial criminal being studied. For example,

- **所有地理学分析法的表现都优于随机方法:** 基于本文的数据库, 相对于随机方法, 任何一个地理学分析法在命中比分上都有显著的提高.
- **所有非随机的地理学分析法都具有大致相同的性能:** 不同的空间模型, 时空模型, 以及整合模型的命中比分间的差异在统计上并不显著, 这表明所有的地理学分析方法都具有相同水平的性能. 这个结论与 Snook 等人得到的相似, Snook 等人的结论表明复杂的地理学分析方法并不比简单的地理学分析方法更为精确 [4].
- **特定的地理学分析方法并不适用于所有的连环犯罪:** 我们可以看到表1中, 本文最佳的单一模型的精确程度很大适度上取决于具体所研究的连环犯罪. 例如, 本文最佳的单一模型即时间指数衰减方

微信搜一搜 数学模型



(a) Locations of Peter Sutcliffe murders. Map generated via Google Static Maps API [20].



(b) Temporal exponential decay model applied to Peter Sutcliffe murders.

Figure 8: Comparison of the prioritized search area produced by the temporal exponential decay model to the map of the geographical area of murders committed by Peter Sutcliffe. This prioritized search area suggests that the Bradford area is at the highest risk of attack.

微信搜一搜 数学模型

our best individual method, the temporal exponential decay method, is an accurate predictor of the D.C. sniper attacks (with an accuracy of 0.056), but is grossly inaccurate when applied to the serial rapist in Columbus, OH (with an accuracy of 0.533, which is worse than the random method).

法, 在华盛顿连环狙击案中作出了准确的预测 (精度为 0.056), 但在俄亥俄州哥伦布市的连环强奸案中却非常不准确 (精度为 0.533, 这比随机方还差)

8 Future Work / 模型展望

- **Standardization of data set:** It is currently difficult to compare geographical profiling method performance to previous research because each study uses a different set of serial crime data to determine performance results. An effort to compile a large, comprehensive list of serial crime data to which all methods would be compared would greatly help the development of geographical profiling techniques.
- **Use of other relevant geographical information:** In this paper we only consider the spatiotemporal relationships between crimes. Work by O'Leary and Brown et al. develop the notion of a feature space that groups crime locations by their relevant geographical features, such as population density or proximity to a major highway [12, 11]. The probability that a future crime happens at a particular location is then determined by the proximity of the location to previous crimes in the feature space. Having more information about what connects crime locations could potentially improve our geographical profiling methods.
- **Use of relevant information about the criminal:** Research suggests that characteristics of the criminal, such as gender, race, and age, play a role in determining their probable spatial behavior [9, 5, 10]. By taking these factors into account, we may be able to improve the accuracy of our models for a specific criminal.
- **Evaluation of cost to law enforcement:** Law enforcement agencies typically purchase computer-generated geographical profiling information [4, 7]. Research by Snook et al. suggests that a person with minimal training in geographical profiling techniques can determine

- **标准化的数据库:** 目前, 比较地理学分析方法与以前的研究工作的性能还很难, 这是因为每份研究使用了不同的连环犯罪数据来确定其性能. 建关一个大型的, 全面的连环犯罪的数据库, 使得所有的方法都可以相互比较, 这将有助于地理学分析方法的发展.
- **使用其它相关的地理信息:** 在本文中, 我们仅考虑了连环犯罪子案件间的时间和空间关系. O'Leary 和 Brown 等人的研究其于相关的地理特征 (比如人口密度, 临近的主要公路) 提出了作案地点的空间特征的概念 [12, 11]. 某个地点未来的作案发生的概率将根据与以前作案的空间特征相似程度来确定. 拥用更多与作案地点有关连的信息会潜在提高我们的地理学分析方法的性能.

- **使用犯罪分子的相关信息:** 研究表明犯罪分子的特征, 比如性别, 种族, 以及年龄, 在确定他们的犯罪行为上发挥了一定的作用. 针对某个特定的案件, 通过考虑这些因素, 我们也许能够提高我们模型的精确程度.
- **评估执法成本:** 警方通常购买由计算机生成的地理学分析信息 [4, 7]. Snook 等人的研究表明一个在地理学分析技术方面欠培训的人员能够判断连环犯罪的作案者的可能居住地, 其精度与一个复杂

微信搜一搜 数学模型

the probable residence of a serial criminal with just as much accuracy as a complex computer-generated geographical profiling method [8]. Further research should be done to determine if the cost of proprietary geographical profiling software is worth the quality of the information provided to law enforcement agencies.

的计算机生成的地理学分析方法相当 [8]. 应该作进一步的研究来确定相对于提供给警方的信息的质量, 独有的地理分析软件的成本是值得的.

微信公众号：数学模型
微信号：MATHmodels
微信搜一搜
数学模型

References

[1] Ronald M. Holmes and Stephen T. Holmes. *Serial Murder*. SAGE Publications, Thousand Oaks, California, 1998.

[2] Donald Brown and Justin Stile. Geographic profiling with event prediction. *Systems, Man and Cybernetics, 2003. IEEE International Conference on*, 4:3712–3719, 2003.

[3] Scotia J. Hicks and Bruce D. Sales. *Criminal Profiling: Developing an Effective Science and Practice*. American Psychological Association, Washington, DC, 2006.

[4] Brent Snook, Michele Zito, Craig Bennell, and Paul J. Taylor. On the complexity and accuracy of geographic profiling strategies. *Journal of Quantitative Criminology*, 21(1):1–26, 2005.

[5] Brent Snook, Richard M. Cullen, Andreas Mokros, and Stephan Harbort. Serial murderers' spatial descisions: Factors that influence crime location choice. *Journal of Investigative Psychology and Offender Profiling*, 2(3):147–164, 2005.

[6] D. Kim Rossmo. Place, space, and police investigations: Hunting serial violent criminals. In *Crime and Place: Crime Prevention Studies*. Willow Tree Press, NY, 1995.

[7] D. Kim Rossmo. Geographic profiling: Target patterns of serial murderers. *Unpublished doctoral dissertation*, 1995.

[8] Brent Snook, Paul J. Taylor, and Craig Bennell. Shortcuts to geographic profilling success: A reply to Rossmo (2005). *Applied Cognitive Psychology*, 19(5):655–661, 2005.

[9] Jasper J. van der Kemp and Peter J. van Koppen. *Fine-Tuning Geographical Profiling*, chapter 17. Criminal Profiling. Humana Press, 2007.

[10] Eric Beauregard, Jean Proulx, and D. Kim Rossmo. Spatial patterns of sex offenders: Theoretical, empirical, and practical issues. *Aggression and Violent Behavior*, 10(5):579–603, 2005.

[11] Donald Brown and Hua Liu. Spatial-temporal event prediction: A new model. *Systems, Man, and Cybernetics, 1998. IEEE International Conference on*, 3:2933–2937, 1998.

[12] Mike O'Leary. The mathematics of geographic profiling. *Journal of Investigative Psychology and Offender Profiling*, 6(3):253–265, 2009.

[13] Spotcrime.com. `http://www.spotcrime.com`. Accessed on February 19, 2010.

微信搜一搜 数学模型

[14] D. Kim Rossmo. Geographic heuristics or shortcuts to failure?: Response to Snook et al. *Applied Cognitive Psychology*, 19(5):651–654, 2005.

[15] Steven Gottlieb, Sheldon Arenberg, and Raj Singh. *Crime Analysis: From First Report to Final Arrest*. Alpha, 1994.

[16] Joshua Kent and Michael Leitner. Efficacy of standard deviational ellipses in the application of criminal geographic profiling. 4(3):147–165, 2007.

[17] Robert T. Clemen and Robert L. Winkler. Aggregating probability distributions. In *Advances in Decision Analysis*. Cambridge University Press, Cambridge, UK, 2007.

[18] Averill M. Law. *Simulation Modeling and Analysis*. McGraw-Hill, 3rd edition, 2000.

[19] Nicole Ward Jouve. *'The Streetcleaner' The Yorkshire Ripper Case on Trial*. Marion Boyars Publishers, London, 1986.

[20] Google Static Maps API. `http://code.google.com/apis/maps/documentation/staticmaps/`. Accessed on February 22, 2010.

微信公众号：数学模型 微信号：MATHmodels

微信搜一搜 数学模型