**一种基于自激点过程的犯罪协同进化方法**

**技术领域**

本发明涉及特征选择和基于自激点过程的犯罪建模领域，尤其涉及基于犯罪事件的协同进化建模方法。

**背景技术**

犯罪建模作为解决日益增加的犯罪率难题的一种科学的方法，试图通过社会统计学以及犯罪心理学等多领域知识，从中挖掘出犯罪事件发生的内在规律并且尝试去预测未来可能的犯罪发生地点。随着社会经济的不断发展，犯罪形式与手段复杂多变，然而犯罪事件始终在一定的空间范围内发生。以空间为视角的犯罪地理研究已经成为犯罪研究中非常重要的方面。另一方面犯罪学从不同的角度研究关于犯罪行为的随机性与犯罪类型的多样性。

现有的犯罪建模方法可以归纳为两种主要的方式。第一种方法是基于地理信息系统的图像与数据结合的方式建模犯罪，这些工作侧重于从风险地形建模、回归分析等统计分析方法对城市内部犯罪绘制风险地图。主要研究城市的内部犯罪，所以有很多工作做的是基于时空模式检测、或者结合城市POI等动态扩散技术。然而，这些工作都只是把历史数据作为静态参考，而没有考虑到一个犯罪事件发生之前的历史犯罪事件对当前时刻的影响。

第二种方法是建模成犯罪网络来分析问题，一般研究的对象是犯罪分子或者受害者这种实体，针对一些具体的犯罪场景比如毒品交易网络，来研究犯罪组织内部的结构序列模式来挖掘用户与物品之间的关系。然而，这些工作对数据的依赖性非常强，并且特征之间的交互性、可解释性都会丢失，这种类型的工作包括：基于图论或者仿真进行犯罪建模；根据犯罪特征的回归方程或者时空分布拟合来解决犯罪预测。他们的模型都做了很强的参数化假设，这样就会导致丢失犯罪事件与不同犯罪类型和犯罪地点之间的动态交互性。

现有的犯罪建模方法虽然可以针对具体某个城市内部做犯罪预测，但是存在以下两点不足：第一，忽略了犯罪与地点之间的动态相互影响；第二，忽略了异构数据的大规模性、不确定性和模糊性。犯罪特征的动态交互性可以保证一个犯罪事件发生前后的环境信息不丢失。因此可以用一个自激点过程模型来描述犯罪类型与地点之间相互作用模式的可靠性和适用性，以发现具有不同潜在特征的犯罪和地点的显著性和持久性特征。

**发明内容**

提出一种基于自激点过程的犯罪协同进化方法。本发明利用以下原理：协同进化模型可以构建犯罪事件和犯罪地点、时间、类型的交互性。协同进化模型从模糊特征集合中选择出核心子集，利用犯罪地点与犯罪类型的 embedding 过程，保留犯罪特征的彼此交互性；基于自激点过程机制可以解决历史数据对当前犯罪发展的影响，考虑特定地区的基准犯罪数据与相邻区域对犯罪强度的影响，建模出人口流动规模较大的地区之间的相互影响。使用高斯核函数触发固定时间窗口下犯罪事件的发生，能够有效解决由于人口流动以及时间推移带来的犯罪地点预测不准确的问题。

为实现上述任务，本发明采用以下技术方案：

一种基于自激点过程的犯罪协同进化方法，包括以下步骤：

步骤1：预处理犯罪数据集，将中国裁判文书网最近5年内成都地区的各种犯罪类型的犯罪事件提取，并获取成都市259个行政地区与当地的出租车轨迹数据。

步骤2：使用模糊集合的特征选择方法提取犯罪核心特征，用一个来表示一个犯罪信息系统，其中表示犯罪实例的集合，表示一个非空的有限的特征集合。定义犯罪特征集之间的运算：

步骤3：从特征集合中找出核心特征子集，其中来表示对的依赖程度，其中，都是的子集。则核心特征子集需满足下述条件：

步骤4：利用挑选出来的核心特征子集，将犯罪数据映射为二分图，将犯罪地点作为一部分，犯罪类型作为另一部分。假设现在给定种犯罪类型和个犯罪地点，现在有个犯罪事件列表的记录为：表示存在条犯罪记录，其中代表一个犯罪事件，表示犯罪类型，是犯罪地点，是该犯罪事件发生的时间，而是步骤3中基于模糊集合的在同一犯罪类型和地点之间的交互得到的核心特征向量。

步骤5：对犯罪类型和犯罪地点分别做embedding，对犯罪类型的update过程，形式上可以描述为：

对犯罪地点的update过程，形式上可以描述为：

其中表示交互特征；表达模型的自我进化性，当前的特征受到历史特征的影响，同时考虑不同犯罪类型和地点的内在演变性；而与表达了犯罪类型与犯罪地点的彼此作用效果；是基于同一犯罪类型，第次犯罪与第次犯罪的时移。

步骤6：自激点过程种犯罪强度值表示某一特定犯罪类型在事件和犯罪地点的犯罪率。强度值取决于时间与给定的时间窗口内的历史犯罪事件记录和犯罪区域内的连接数，形式上，强度函数可以描述为：

其中，，表示特定时间窗口内的所有犯罪类型的交互，是与地区直接相邻的区域；，建模出人口流动规模较大的跨区域之间的影响，是交通流量的规模；取决于该犯罪类型在地区的历史犯罪记录。

步骤7：使用触发函数来表征犯罪事件发生的交互过程，形式上可以描述为：

这里，表示时间戳之前的时间窗口长度，是该时间窗口内的犯罪事件数量，是第次犯罪的时间戳，确定了历史犯罪的平均时间影响，可以表示历史记录在犯罪地点的影响随着时间间隔的增加而逐渐衰减；触发函数种表示公共交通流量，决定了衰减速度，从而建模出犯罪活动的动态过程。

**附图说明**

图1为本发明实例中基于协同进化模型的犯罪建模方法流程图。图2表示自激点过程在成都地区的犯罪交互过程演示图。

**具体实施方式**

以下结合附图，对本发明上述的和另外的技术特征和优点做详细的说明。

本发明共有2幅附图，请参阅图1所示，本发明的具体步骤如下：

步骤1：预处理犯罪数据集，将中国裁判文书网最近5年内成都地区的各种犯罪类型的犯罪事件提取，并获取成都市259个行政地区与当地的出租车轨迹数据。

步骤2：使用模糊集合的特征选择方法提取犯罪核心特征，用一个来表示一个犯罪信息系统，其中表示犯罪实例的集合，表示一个非空的有限的特征集合。定义犯罪特征集之间的运算：

步骤3：从特征集合中找出核心特征子集，其中来表示对的依赖程度，其中，都是的子集。则核心特征子集需满足下述条件：

步骤4：利用挑选出来的核心特征子集，将犯罪数据映射为二分图，将犯罪地点作为一部分，犯罪类型作为另一部分。假设现在给定种犯罪类型和个犯罪地点，现在有个犯罪事件列表的记录为：表示存在条犯罪记录，其中代表一个犯罪事件，表示犯罪类型，是犯罪地点，是该犯罪事件发生的时间，而是步骤3中基于模糊集合的在同一犯罪类型和地点之间的交互得到的核心特征向量。

步骤5：对犯罪类型和犯罪地点分别做embedding，对犯罪类型的update过程，形式上可以描述为：

对犯罪地点的update过程，形式上可以描述为：

其中表示交互特征；表达模型的自我进化性，当前的特征受到历史特征的影响，同时考虑不同犯罪类型和地点的内在演变性；而与表达了犯罪类型与犯罪地点的彼此作用效果；是基于同一犯罪类型，第次犯罪与第次犯罪的时移。

步骤6：自激点过程种犯罪强度值表示某一特定犯罪类型在事件和犯罪地点的犯罪率。强度值取决于时间与给定的时间窗口内的历史犯罪事件记录和犯罪区域内的连接数，形式上，强度函数可以描述为：

其中，，表示特定时间窗口内的所有犯罪类型的交互，是与地区直接相邻的区域；，建模出人口流动规模较大的跨区域之间的影响，是交通流量的规模；取决于该犯罪类型在地区的历史犯罪记录。

步骤7：使用触发函数来表征犯罪事件发生的交互过程，形式上可以描述为：

这里，表示时间戳之前的时间窗口长度，是该时间窗口内的犯罪事件数量，是第次犯罪的时间戳，确定了历史犯罪的平均时间影响，可以表示历史记录在犯罪地点的影响随着时间间隔的增加而逐渐衰减；触发函数种表示公共交通流量，决定了衰减速度，从而建模出犯罪活动的动态过程。

本发明为基于自激点过程的犯罪协同进化方法，提出了一个自激点过程模型来描述犯罪类型与地点之间相互作用模式的可靠性和适用性，并发展了一种基于模糊粗糙集的特征选择方法，以发现具有不同潜在特征的犯罪和地点的显著性和持久性特征。

以上所述仅为本发明的较佳实施例，对于发明而言仅仅是说明性的，而非限制性的。本专业技术人员理解，在发明权利要求所限定的精神和范围内可对其进行许多改变、修改等，但都将落入本发明的保护范围内。

一、申请发明专利或者实用新型专利应当提交说明书，一式一份。

二、说明书应当打字或者印刷，字迹应当整齐清晰，呈黑色，符合制版要求，不得涂改，字高在3.5毫米至4.5毫米之间，行距在2.5毫米至3.5毫米之间。说明书首页用此页，续页可使用同样大小和质量相当的白纸。纸张应当纵向使用，只限使用正面，四周应当留有页边距：左侧和顶部各25毫米，右侧和底部各15毫米。

三、说明书第一页第一行应当写明发明创造名称，该名称应当与请求书中的名称一致，并左右居中。发明创造名称与说明书正文之间应当空一行。说明书格式上应当包括下列五个部分，并且在每一部分前面写明标题：

技术领域

背景技术

发明内容

附图说明

具体实施方式

说明书无附图的，说明书文字部分不包括附图说明及其相应的标题。说明书文字部分可以有化学式、数学式或者表格，但不得有插图。

四、涉及核苷酸或氨基酸的申请，应当将该序列表作为说明书的一个单独部分，并单独编写页码。申请人应当在申请的同时提交与该序列表相一致的光盘或软盘，该光盘或软盘应符合国家知识产权局的有关规定。

五、说明书应当在每页下框线居中位置顺序编写页码。