**一种基于动态犯罪网络的事件场景建模方法**

**技术领域**

本发明涉及多源数据特征提取与基于动态犯罪网络的场景建模领域，尤其涉及基于犯罪事件的动态网络建模的方法。

**背景技术**

现有的对犯罪事件的场景建模方法可以归纳为两种主要的方式。第一种方法是基于时空犯罪热点数据来从时间、空间特征维度来对犯罪事件进行抽象和建模，这些工作可以利用 GIS 技术获得时间序列模型，或者通过数学方法搭建密度转移模型。这种方法增加了主体对环境信息的反馈功能，与以往的模型相比，反馈功能的实现有利于实现系统的动态性和预测的可靠性。虽然争对某一个城市效果较好，但忽视了犯罪活动的空间扩散效应，同时对数据要求严格。

第二种方法是基于犯罪网络结构之间的关系从而对犯罪活动中的各个参与者进行抽象建模。一般研究的对象是犯罪分子或者受害者这种实体，有基于代理技术，基于图论或者仿真等技术。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。通过研究在外界破坏了网络中的节点之后网络的承受能力和恢复能力。利用不同的破坏策略和恢复策略对网络进行攻击，从而揭示犯罪网络在破坏和复原之间的动态过程。

现有的犯罪场景建模方法虽然可以利用提取的犯罪特征融合犯罪主体与环境的交互性来建模犯罪网络，但是存在以下两点不足：第一，在数据层面，尽管将一堆异构数据丢入到深度学习模型中，或许能够得到不错的预测效果，但是特征之间的交互性、可解释性都会丢失，并且换一个数据集，其结果将不确定。同时在分析大规模异构数据的时候如何降低特征的高维性也是一个难题。第二，在方法层面，用提取的特征做回归分析或者在时空维度上预测犯罪率，现有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，并且依赖于当前静态的历史数据，而忽略了一个犯罪事件在动态变化的犯罪场景中的多变性。

**发明内容**

提出一种基于动态犯罪网络的事件场景建模方法。本发明利用以下原理：利用一种新的时空文本数据统计建模框架，演示了犯罪事件在动态网络生成中的应用，同时利用提取的犯罪特征融合犯罪主体与环境的交互性来建模犯罪活动；马尔可夫链与空间自相关技术相结合，用来研究犯罪事件发生后的无后效条件下时间和状态均为离散的随机转移过程,并结合空间权重矩阵, 建立了犯罪区域间的空间交互关系。在犯罪网络中建立模型关联集中在犯罪实体之间的关系上，将节点中包含的个人信息归纳到犯罪网络的分析中。并且将一个犯罪活动抽象为交易利益最大化的模型，最终将一个犯罪事件嵌入到复杂的动态犯罪网络中。

为实现上述任务，本发明采用以下技术方案：

一种基于动态犯罪网络的事件场景建模方法，包括以下步骤：

步骤一：预处理宝贝回家网、中国裁判文书网数据，提取出人口统计数据、教育水平、人口流动性这些犯罪学特征，并过滤掉犯罪事件中缺乏受害人、犯罪分子、犯罪事件发生地点以及法律判罚元素的数据项。

步骤二：对犯罪事件中的空间特征组进行处理。利用地理集中度系数来衡量犯罪集聚程度的变化，而犯罪流动性，表示从拐出区域拐向拐入区域的实际人口迁移数与期望数之比。

步骤三：根据空间马尔科夫链与空间自相关建立在无后效条件下，犯罪区域之间在时间和状态均为离散的随机转移的相关性：将各省域的儿童拐卖数据离散化为个等级，建立一个用于存储年份的种类型的拐卖转移概率。然后结合所有类型，形成一个的转移矩阵。其中矩阵中每个元素的转移概率表示一个空间单元的儿童拐卖数量在年份属于类型在年份转变为类型的概率。

步骤四：根据地理加权回归函数分析驱动各省域的因素，其中是第个样本空间单元的空间位置，表示第个样本点的常数估计值，是连续函数在样本空间单元的值，为第个样本点的独立变量，为误差修正项。

步骤五：犯罪主体间交互建模主要使用效用函数将犯罪网络节点中包含的个人信息归纳到犯罪网络的分析中。并且将一个犯罪活动抽象为交易利益最大化的模型，因此效用函数形式上可以描述为：

其中乘以是组织这次犯罪活动的预期收入，乘以是贿赂的预期成本，用于收买犯罪成员的沉默，防止计划犯罪的信息泄露，是从收益中支付给这些犯罪参与者的薪资。

步骤六：将这个效用函数应用到犯罪网络，其中是节点的集合，也就是个体嫌疑，是连接犯罪个体的弧的集合，得到更一般的形式描述为：

其中表示概率，表示犯罪网络中愿意参与到这次犯罪活动中的个体，其中每一个个体都是整个犯罪网络节点集合中的元素，能力越大，越容易参与到犯罪活动中。

步骤七：由步骤6得到的效用函数三部分分别进行表示：

其中，表示实施犯罪计划的概率。是给定链接的泄密概率, 是总佣金，是成本因子，和前边的类似，表示每个犯罪网络成员作为其犯罪能力函数的支出。

步骤八：利用效用函数的最大值确定下界：，接着确定了应该给这些犯罪参与者个体支付的最大单位费用： 。

步骤九：求出在约束条件下的最优解，其中效用函数的最大值可以描述为：

而模型的约束条件可以表示为：

最终得到从（犯罪组织者）到（犯罪个体参与者）之间最佳关联的集合，上述的约束条件保证了能建模出犯罪网络中的一条最佳路径。

本发明的有益效果是：由于被拐儿童之间一般是彼此不认识的，并且他们有独立的生活状态。例如一个被拐儿童在被拐前处于这种生活环境中，此使他的状态是。当他在做的时候被拐卖，此时他的身份别为了受害者，也就是。对于犯罪分子，通过调研一些法律裁判文件，发现大部分拐卖儿童犯罪网络都存在上下级关系，比如雇佣关系，或者存在中间人介绍这种层级结构。那么此时的被拐儿童和犯罪者就相当于一种交易。既然是交易就存在着供需关系，存在着安全性和经济收益等问题。由于犯罪网络中的两个节点可以通过各种不同的路径关联，因此这种方法可以很好地识别表示两个节点之间“最佳”关联的路径。同时寻找最佳关联的过程其实也就是一个犯罪网络形成的过程。

**附图说明**

图1为本发明实例中基于动态犯罪网络的事件场景建模方法模型图。

**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1幅附图，请参阅图1所示，本发明的具体步骤如下：

步骤一：预处理宝贝回家网、中国裁判文书网数据，提取出人口统计数据、教育水平、人口流动性这些犯罪学特征，并过滤掉犯罪事件中缺乏受害人、犯罪分子、犯罪事件发生地点以及法律判罚元素的数据项。

步骤二：对犯罪事件中的空间特征组进行处理。利用地理集中度系数来衡量犯罪集聚程度的变化，而犯罪流动性，表示从拐出区域拐向拐入区域的实际人口迁移数与期望数之比。

步骤三：根据空间马尔科夫链与空间自相关建立在无后效条件下，犯罪区域之间在时间和状态均为离散的随机转移的相关性。

步骤四：根据地理加权回归函数分析驱动各省域的因素，其中是第个样本空间单元的空间位置，表示第个样本点的常数估计值，是连续函数在样本空间单元的值，为第个样本点的独立变量，为误差修正项。

步骤五：犯罪主体间交互建模主要使用效用函数将犯罪网络节点中包含的个人信息归纳到犯罪网络的分析中。并且将一个犯罪活动抽象为交易利益最大化的模型，因此效用函数形式上可以描述为：

其中乘以是组织这次犯罪活动的预期收入，乘以是贿赂的预期成本，用于收买犯罪成员的沉默，防止计划犯罪的信息泄露，是从收益中支付给这些犯罪参与者的薪资。

步骤六：将这个效用函数应用到犯罪网络，其中是节点的集合，也就是个体嫌疑，是连接犯罪个体的弧的集合，得到更一般的形式描述为：

其中表示概率，表示犯罪网络中愿意参与到这次犯罪活动中的个体，其中每一个个体都是整个犯罪网络节点集合中的元素，能力越大，越容易参与到犯罪活动中。

步骤七：由步骤6得到的效用函数三部分分别进行表示：

其中，表示实施犯罪计划的概率。是给定链接的泄密概率, 是总佣金，是成本因子，和前边的类似，表示每个犯罪网络成员作为其犯罪能力函数的支出。

步骤八：利用效用函数的最大值确定下界：，接着确定了应该给这些犯罪参与者个体支付的最大单位费用： 。

步骤九：求出在约束条件下的最优解，其中效用函数的最大值可以描述为：

而模型的约束条件可以表示为：

最终得到从（犯罪组织者）到（犯罪个体参与者）之间最佳关联的集合，上述的约束条件保证了能建模出犯罪网络中的一条最佳路径。

本发明为基于动态犯罪网络的事件场景建模方法，通过提取的犯罪特征融合犯罪主体与环境的交互性来建模犯罪网络，且将一个犯罪活动抽象为交易利益最大化的模型，最终将一个犯罪事件嵌入到复杂的动态犯罪网络中。

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。