



(12)发明专利申请

(10)申请公布号 CN 106022296 A

(43)申请公布日 2016. 10. 12

(21)申请号 201610380318.6

(22)申请日 2016.06.01

(71)申请人 银江股份有限公司

地址 310012 浙江省杭州市益乐路223号1幢1层

(72)发明人 蒋伶华 李建元 陈涛 李丹 温晓岳

(74)专利代理机构 杭州斯可睿专利事务所有限公司 33241

代理人 王利强

(51)Int. Cl.

G06K 9/00(2006.01)

G06K 9/62(2006.01)

G06K 9/64(2006.01)

G08G 1/017(2006.01)

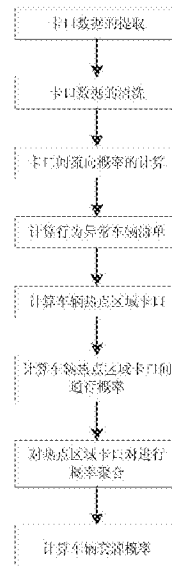
权利要求书2页 说明书15页 附图2页

(54)发明名称

一种基于车辆热点区域的概率聚合的套牌车检测方法

(57)摘要

一种基于车辆热点区域的概率聚合的套牌车检测方法,包括以下步骤:S1.卡口数据的提取;S2.卡口数据的清洗;S3.计算卡口间流向概率;S4.计算行为异常车辆清单:根据S2中提取出来的每一辆车的卡口对,计算卡口对概率,按照概率低于设定阈值Ti的次数,降序排列,根据需要选取前N行作为行为异常车辆清单;S5.计算车辆热点区域卡口点;S6.计算车辆在两热点区域卡口点之间通行的概率;S7.根据S6所得卡口间概率,对所有卡口进行聚合;S8.根据车辆卡口聚合情况计算车辆套牌概率Z.本发明克服随机性的卡口号牌识别错误,利用车辆活动区域相对固定的特性,筛选出套牌嫌疑度较大的车辆、大大缩小嫌疑范围、实用性良好。



1. 一种基于车辆热点区域的概率聚合的套牌车检测方法,其特征在于:所述套牌车辆检测方法包括以下步骤:

S1. 卡口数据的提取:获取卡口过车记录数据,保留需要的维度,包括卡口编号、卡口方向、号牌号码、号牌种类,过车时间;

S2. 卡口数据的清洗;

S3. 计算卡口间流向概率,过程如下:

3.1、根据行车轨迹,提取卡口向量:将卡口数据按照号牌号码,卡口过车时间进行排序,得到每一辆车的行车轨迹,用所经过的卡口序列 $(K_a, K_b, K_c \cdot K_i \cdot K_n \cdot)$ 表示,提取每一辆行车轨迹中相邻的卡口对,形成卡口向量序列 $(K_a, K_b) \cdot (K_i, K_j) \cdot$ 。 K_a ,表示车辆经过的第a个卡口,向量 (K_a, K_b) 表示车辆顺序经过两个卡口 K_a 和 K_b ,且 K_a 和 K_b 在行车轨迹中相邻;

3.2、计算卡口间流向概率:对卡口 K_i 计数,得到 K_i 卡口流出车辆总和,用 $\text{count}(K_i)$ 表示。对向量 (K_i, K_j) 计数,得到卡口 K_i 到卡口 K_j 通行车辆总和,用 $\text{count}(K_i, K_j)$ 表示。卡口 K_i 到卡口 K_j 的流向概率为 $P_{ij} = \text{count}(K_i, K_j) / \text{count}(K_i)$;

S4. 计算行为异常车辆清单:根据S2中提取出来的每一辆车的卡口对,计算卡口对概率,按照概率低于设定阈值 T_i 的次数,降序排列,根据需要选取前N行,作为行为异常车辆清单,N根据实际需要选取;

S5. 计算车辆热点区域卡口点:热点活动区域卡口点指的是车辆经常被捕获的卡口,统计车辆在每个卡口点出现的次数,如果大于设定的阈值 M_v ,则认为该卡口是车辆的热点活动区域卡口点;

S6. 计算车辆在两热点区域卡口点之间通行的概率:对每一辆车在S5中计算的热点卡口做笛卡尔积,并且根据S2计算的概率,给每一对卡口进行赋值,并且选取正反向中较大的概率;如果卡口 K_i 到卡口 K_j 的概率 $P(K_i, K_j) <$ 卡口 K_j 到卡口 K_i 的概率 $P(K_j, K_i)$,那么将 $P(K_j, K_i)$ 赋值给卡口对 (K_i, K_j) ;

S7. 根据S6所得卡口间概率,对所有卡口进行聚合,确保各个分组内部,对于任一卡口对 (K_i, K_j) ,存在一条路径从 K_i 到 K_j ,路径上所有相邻卡口对的概率均不小于设定阈值 T_i ,各个分组之间,对于任一卡口 K_i ,其他组不存在卡口 K_j ,满足 $P(K_i, K_j)$ 大于设定的阈值 T_i ;

S8. 计算车辆套牌的概率,根据车辆卡口聚合情况计算车辆套牌概率 $Z, Z = N * i * s^4$,N表示车辆热点卡口总数,i取决于分组数量L,如果 $L = 2, i = 1$,如果 $L = 3, i = 1/2$,否则 $i = 0$;s取决于分组中卡口的数量比, NA_i 表示每一组中卡口的数量; $s = \max(NA_i, 2) / \max(NA_i, 1)$, $\max(NA_i, 1)$ 表示数量最大的一个, $\max(NA_i, 2)$ 表示数量第二大的。

2. 如权利要求1所述的一种基于车辆热点区域的概率聚合的套牌车检测方法,其特征在于:所述步骤S7中个,聚合过程如下:

Step7.1. 获取可以车辆号牌清单;

Step7.2. 从号牌清单中取出一个号码,根据号码,获取该号牌热点活动区域卡口 $(K_1, K_2, K_3 \cdot \cdot \cdot K_n)$,记为数组A0,并且根据车辆出现在该卡口的次数倒叙排列;

Step7.3. 将A0中的第一个卡口 K_1 添加到数组A1中,并且将 K_1 从A0中去除;

Step7.4. 判断 K_1 与A0中每一个卡口之间的无向概率 $Q(K_1, K_i)$ 是否大于阈值 T_i ,如果是,将 K_i 添加到A1和空数组A_new,并且将 K_i 从数组A0中去除;

Step7.5. 从数组A_new中取出一个卡口 K_j ,判断 K_j 与A0中每一个卡口之间的无向概率 Q

(K_j, K_i)是否大于阈值 T_i ,如果是,将 K_i 添加到 A_1 和空数组 A_{new} ,并且将 K_i 从数组 A_0 中去除,将 K_j 从数组 A_{new} 中去掉;

Step7.6.重复Step7.5,直到 A_{new} 为空;

Step7.7.增加数组 A_i ,重复步骤Step7.3-7.6;

Step7.8.重复Step7.7,直到 A_0 为空;

Step7.9.重复Step7.2-7.8,直到号牌清单为空。

3.如权利要求1或2所述的一种基于车辆热点区域的概率聚合的套牌车检测方法,其特征在于:所述步骤S2中,卡口数据的清洗过程如下:

2.1、清洗重复记录;

2.2、清洗脏数据,所述脏数据是指号牌号码不符合命名规则的记录;

2.3、确定时间区间,清洗该区间内记录不齐全的卡口数据,这些情况包括卡口损坏和断电导致卡口无法工作,或者网络中断和磁盘容量不足导致数据未能存储,清洗该卡口所有数据。

4.如权利要求1或2所述的一种基于车辆热点区域的概率聚合的套牌车检测方法,其特征在于:所述步骤S4中,阈值 T_i 根据卡口号牌识别正确率 F 进行设定,对卡口 K_i 而言, T_i 的取值应满足 $F \leq \sum_j P_{ij} < F - T_i$,其中 j 满足 $P_{ij} > T_i$,卡口号牌识别正确率 $F = 1 - (\text{通过卡口的但是号牌不在号牌库中的过车记录数}) / (\text{所有过车记录数})$, T_i 表示卡口 K_i 的阈值, P_{ij} 表示卡口 P_i 到其他卡口的概率。

一种基于车辆热点区域的概率聚合的套牌车检测方法

技术领域

[0001] 本发明涉及智能交通领域,具体涉及一种套牌车检测方法。

背景技术

[0002] 套牌车俗称克隆车,是指通过伪造或者非法套取其它车辆号牌及行驶证等手续上路行驶的车辆。随着经济的发展,车辆保有量越来越多,套牌车也不断增加。套牌车会扰乱公安机关对公共安全的管控,由于套牌车没有合法手续和保险,一旦发生交通事故,驾驶人极易逃逸。套牌车还为犯罪活动提供作案工具,大大增加了破案难度。套牌车会逃避各种税费,造成国家税费的大量流失,并扰乱运输市场的秩序。套牌车会损坏真车主的合法权益,在车辆交通违法、事故处理等方面,真车主往往要充当“冤大头”。

[0003] 套牌车这一违章行为的识别查处难度很大,民警在执勤过程中很难短时间根据车牌及外观特征人工判断出车辆是否为套牌车。现在已经有一些基于卡口数据的套牌车自动化识别方法。卡口指的是,采用先进的光电、计算机、图像处理、模式识别、远程数据访问等技术,对监控路段的机动车道、非机动车道进行全天候实时监控并记录相关图像数据,并自动获取车辆的通过时间、地点、行驶方向、号牌号码、号牌颜色、车身颜色等数据。

[0004] 已公开的专利及审查中的专利中,主要分成三大类,1、基于图像识别,通过车辆特征进行判断。2、基于辅助装置:例如安装电子车牌,预留安全密码等等。3、基于卡口数据,对车辆轨迹进行分析。

[0005] 基于第一类方法的专利有:

[0006] 一:申请号[201510102368.3]提出的,基于识别车辆的车牌号和品牌信息的套牌车检测方法。1、识别车辆的车牌号和品牌信息,若车牌号和品牌信息不对应,则将该车辆加入套牌车嫌疑库;2、对加入套牌车嫌疑库的车辆进行精确判别,去除误报的车辆;3)统计车辆在套牌车嫌疑库中出现的次数,当某车辆的出现次数超过设定值K时发出警报。

[0007] 二:申请号[201410333789.2]提出的,一种基于车辆身份识别的套牌车识别的方法与装置。1、从图片或视频中提取出车辆,识别该车辆特征。2、为与交管所数据库进行比对,当发现车辆某些特征与数据库内特征不匹配时,进行报警。2、在市区内卡口进行稽查布控,当发现特征相同的车辆在同一时间的不同位置出现时,则进行套报警。3、历史数据进行比对,当发现车牌号码相同的车辆,车辆其他特征不相同,进行报警。

[0008] 以上专利方法中,天气、光线等因素会干扰图像的识别,部分国产车和进口车外形较为相似,也会给识别造成困难。存在原车和套牌车品牌、型号相同则无法检测等缺陷。

[0009] 基于第二类方法的专利有:

[0010] 一:申请号[201320365827.3]提出的,一种基于超高频射频识别技术的假、套牌车辆识别装置。基于超高频射频识别技术的假、套牌车辆识别装置,通过射频识别读写设备对汽车上的电子标签进行检测,并对电子标签内的信息进行判断,然后与车牌图像采集设备得出的车牌号码进行比对,进行套牌车判别。

[0011] 二:申请号[200910107671.7]提出的,一种基于安全密码的机动车套牌假牌违法

行为检测方法。交警内部车辆管理系统平台预留车辆安全检测密码；执法交警手持远程控制终端输入被检查车辆区域代码和车牌号码；与注册信息对比；输入被查车辆安全检测密码；输入密码与预留密码自动对比，判断是否套牌假牌。

[0012] 以上专利需要涉及到额外的平台建设和车辆改造。

[0013] 基于第三类方法的专利有：

[0014] 一：申请号[201210438702.9]提出的，一种基于时间交集的套牌车检测方法。获取车辆进入和离开该位置的时刻以及在该位置的停留时间段，两两配对，如果对应的停留时间段有交集，则存在套牌车，如果无交集，则根据两点之间的理论最短时间和实际时间差来判断是否存在套牌车。该方案中，理论最短时间计算复杂，涉及到路网结构，路段长度，交通拥堵度等诸多参数。

[0015] 二：申请号[201310730531.1]提出的，一种套牌车车辆的区域自动捕获方法。在封闭的区域边界设置监控点，如果车辆从该区域驶出，但是没有驶入信息，则判定该车辆为套牌车辆。该方案中的封闭区域在实际中难以实现。

[0016] 三：申请号[201310034242.8]提出的，一种基于道路网格化的套牌车车辆的捕获方法。将城市道路区域依据网格切分原则进行网格化，检验同一车牌号车辆按时间序列的网格轨迹是否连续，若出现不连续的网格轨迹，则将该网格轨迹对应的车牌号作为疑似套牌车车牌号码。该方案需要详细的路网结构。

[0017] 四：申请号[201410094882.2]提出的，一种基于卡口连通性的套牌分析的方。1、根据车辆经过卡口的顺序是否和预先保存的卡口连通信息一致，如果不一致，再根据预先保存的卡口连通信息确定任一不一致的卡口到其相邻卡口的所有可达路径，判断所述可达路径中是否存在一条路径，该条路径经过的卡口数量小于等于预设值N，如果是，则所述车牌未套牌，否则，所述车牌存在套牌。

[0018] 现有的方法均是基于单次车辆行驶轨迹进行判断。实际情况中，由于卡口号牌识别错误的存在，以上算法得出的套牌车嫌疑清单往往达到几万甚至几十万。

[0019] 随机性的卡口号牌识别错误指的是由于光照、角度等因素，车牌在部分卡口或者时间段能够正确识别，部分卡口或者时间段识别成其他车牌。这种情况占号牌识别错误的绝大部分。非随机性号牌识别错误主要是由于车辆号牌污损引起的，该车牌会被所有卡口识别成固定的错误号牌。例车牌号“浙A****R”，其中字符R右下角被遮挡，但是其余部分均很清晰，所有卡口均将该车牌识别成了“浙A****P”。

发明内容

[0020] 为了克服已有套牌车辆检测方法的由于存在随机性的卡口号牌识别错误、嫌疑范围过大、实用性较差的不足，本发明提供一种克服随机性的卡口号牌识别错误，利用车辆活动区域相对固定的特性，筛选出套牌嫌疑度较大的车辆、大大缩小嫌疑范围、实用性良好的基于车辆活动热点区域的概率聚合的套牌车检测方法。

[0021] 本发明解决其技术问题所采用的技术方案是：

[0022] 一种基于车辆热点区域的概率聚合的套牌车检测方法，所述套牌车辆检测方法包括以下步骤：

[0023] S1. 卡口数据的提取：获取卡口过车记录数据，保留需要的维度，包括卡口编号、卡

口方向、号牌号码、号牌种类,过车时间;

[0024] S2.卡口数据的清洗;

[0025] S3.计算卡口间流向概率,过程如下:

[0026] 3.1、根据行车轨迹,提取卡口向量:将卡口数据按照号牌号码,卡口过车时间进行排序,得到每一辆车的行车轨迹,用所经过的卡口序列 $(K_a, K_b, K_c \cdot K_i \cdot K_n \cdot)$ 表示,提取每一辆行车轨迹中相邻的卡口对,形成卡口向量序列 $(K_a, K_b) \cdot (K_i, K_j) \cdot \dots$,表示车辆经过的第 a 个卡口,向量 (K_a, K_b) 表示车辆顺序经过两个卡口 K_a 和 K_b ,且 K_a 和 K_b 在行车轨迹中相邻;

[0027] 3.2、计算卡口间流向概率:对卡口 K_i 计数,得到 K_i 卡口流出车辆总和,用 $\text{count}(K_i)$ 表示。对向量 (K_i, K_j) 计数,得到卡口 K_i 到卡口 K_j 通行车辆总和,用 $\text{count}(K_i, K_j)$ 表示。卡口 K_i 到卡口 K_j 的流向概率为 $P_{ij} = \text{count}(K_i, K_j) / \text{count}(K_i)$;

[0028] S4.计算行为异常车辆清单:根据S2中提取出来的每一辆车的卡口对,计算卡口对概率,按照概率低于设定阈值 T_i 的次数,降序排列,根据需要选取前 N 行,作为行为异常车辆清单, N 根据实际需要选取;

[0029] S5.计算车辆热点区域卡口点:热点活动区域卡口点指的是车辆经常被捕获的卡口,统计车辆在每个卡口点出现的次数,如果大于设定的阈值 M_v ,则认为该卡口是车辆的热点活动区域卡口点;

[0030] S6.计算车辆在两热点区域卡口点之间通行的概率:对每一辆车在S5中计算的热点卡口做笛卡尔积,并且根据S2计算的概率,给每一对卡口进行赋值,并且选取正反向中较大的概率;如果卡口 K_i 到卡口 K_j 的概率 $P(K_i, K_j) <$ 卡口 K_j 到卡口 K_i 的概率 $P(K_j, K_i)$,那么将 $P(K_j, K_i)$ 赋值给卡口对 (K_i, K_j) ;

[0031] S7.根据S6所得卡口间概率,对所有卡口进行聚合,确保各个分组内部,对于任一卡口对 (K_i, K_j) ,存在一条路径从 K_i 到 K_j ,路径上所有相邻卡口对的概率均不小于设定阈值 T_i ,各个分组之间,对于任一卡口 K_i ,其他组不存在卡口 K_j ,满足 $P(K_i, K_j)$ 大于设定的阈值 T_i ;

[0032] S8.计算车辆套牌的概率,根据车辆卡口聚合情况计算车辆套牌概率 Z , $Z = N * i * s^4$, N 表示车辆热点卡口总数, i 取决于分组数量 L ,如果 $L = 2$, $i = 1$,如果 $L = 3$, $i = 1/2$,否则 $i = 0$; s 取决于分组中卡口的数量比, NA_i 表示每一组中卡口的数量; $s = \max(NA_i, 2) / \max(NA_i, 1)$, $\max(NA_i, 1)$ 表示数量最大的一个, $\max(NA_i, 2)$ 表示数量第二大的。

[0033] 进一步,所述步骤S7中个,聚合过程如下:

[0034] Step7.1.获取可以车辆号牌清单;

[0035] Step7.2.从号牌清单中取出一个号码,根据号码,获取该号牌热点活动区域卡口 $(K_1, K_2, K_3 \cdot \dots \cdot K_n)$,记为数组 A_0 ,并且根据车辆出现在该卡口的次数倒叙排列;

[0036] Step7.3.将 A_0 中的第一个卡口 K_1 添加到数组 A_1 中,并且将 K_1 从 A_0 中去除;

[0037] Step7.4.判断 K_1 与 A_0 中每一个卡口之间的无向概率 $Q(K_1, K_i)$ 是否大于阈值 T_i ,如果是,将 K_i 添加到 A_1 和空数组 A_{new} ,并且将 K_i 从数组 A_0 中去除;

[0038] Step7.5.从数组 A_{new} 中取出一个卡口 K_j ,判断 K_j 与 A_0 中每一个卡口之间的无向概率 $Q(K_j, K_i)$ 是否大于阈值 T_i ,如果是,将 K_i 添加到 A_1 和空数组 A_{new} ,并且将 K_i 从数组 A_0 中去除,将 K_j 从数组 A_{new} 中去掉;

[0039] Step7.6.重复Step7.5,直到 A_{new} 为空;

- [0040] Step7.7.增加数组 A_i ,重复步骤Step7.3-7.6;
- [0041] Step7.8.重复Step7.7,直到 A_0 为空;
- [0042] Step7.9.重复Step7.2-7.8,直到号牌清单为空。
- [0043] 再进一步,所述步骤S2中,卡口数据的清洗过程如下:
- [0044] 2.1、清洗重复记录;
- [0045] 2.2、清洗脏数据,所述脏数据是指号牌号码不符合命名规则的记录;
- [0046] 2.3、确定时间区间,清洗该区间内记录不齐全的卡口数据,这些情况包括卡口损坏和断电导致卡口无法工作,或者网络中断和磁盘容量不足导致数据未能存储,清洗该卡口所有数据。
- [0047] 更进一步,所述步骤S4中,阈值 T_i 根据卡口号牌识别正确率 F 进行设定,对卡口 K_i 而言, T_i 的取值应满足 $F < \sum_j P_{ij} < F - T_i$,其中 j 满足 $P_{ij} > T_i$,卡口号牌识别正确率 $F = 1 - (\text{通过卡口的但是号牌不在号牌库中的过车记录数}) / (\text{所有过车记录数})$, T_i 表示卡口 K_i 的阈值, P_{ij} 表示卡口 P_i 到其他卡口的概率。
- [0048] 本发明的技术构思为:在基于通行概率检测车辆套牌方法为基础,利用车辆热点活动区域特性,缩小排查范围。
- [0049] 实际使用中,由于光线、角度、号牌污损等因素,卡口对于号牌的识别率无法达到100%(一般在96%-98%左右)。车牌在大部分卡口或者时间段能够正确识别,小部分卡口或者时间段识别成其他车牌,识别错误的卡口点往往分布是随机的,且通行次数较少。同时,正常家庭用车基本上会行驶在比较固定的路线上,识别正确的卡口点往往分布在这些固定路线上,且次数较多。因此去掉通过次数较少的卡口点,可以将大部分识别错误的卡口进行排除。
- [0050] 车辆经常通过的卡口往往在空间上具有连续性。如果这些卡口点明显分布呈现两个社区,且社区间关联很弱,说明很有可能是两辆车分别在两个区域活动,也就是说明该车的套牌可能性较高。
- [0051] 本发明的有益效果主要表现在:能够在一定程度上克服卡口随机性的号牌识别错误,极大的缩小排查范围、实用性良好;无需依赖路网结构,适用性较强。

附图说明

- [0052] 图1是基于车辆活动热点区域的概率聚合的套牌车检测方法的流程图。
- [0053] 图2是正常的车辆热点卡口分布,其特点是卡口之间具有较强的关联性。
- [0054] 图3是疑似套牌车辆热点卡口分布,其特点是明显分布成两个社区。
- [0055] 图4是较易识别错误的车辆热点卡口分布,其特点是分布成多个社区。

具体实施方式

- [0056] 下面结合附图对本发明作进一步描述。
- [0057] 参照图1~图4,一种基于车辆活动热点区域的概率聚合的套牌车检测方法,包括如下步骤:
- [0058] S1.数据的提取:获取卡口过车记录数据,保留需要的维度,包括卡口编号、卡口方向、号牌号码、号牌种类,过车时间。(卡口编号唯一确定卡口路口位置,卡口方向确定卡口

拍摄方向,即行车方向,号牌种类和号牌号码唯一确定一辆机动车)

[0059] 本实施例抽取了杭州市1月1日-1月27日489个卡口记录的数据,总共129534497条,卡口数据格式如下表1:

[0060]

字段	数据类型	含义
KKID	VARCHAR(20)	卡口ID
FXBH	VARCHAR(2)	卡口方向
HPHM	VARCHAR(10)	号牌号码
HPLX	VARCHAR(2)	号牌种类
JGSJ	VARCHAR(20)	过车时间

[0061] 表1

[0062] 其中KKID+FXBH唯一确定一个卡口,HPHM+HPZL唯一确定一辆汽车。JGSJ精确到秒,(以下步骤中,卡口ID包含了卡口方向,号牌号码包含了号牌种类,不再赘述)

[0063] S2.数据的清洗,过程如下:

[0064] 2.1、清洗重复数据:对于卡口,捕捉一辆经过的车辆时,可能会产生多条数据,造成数据的重复,如下表2所示:

[0065]

序号	号牌号码	卡口ID	过车时间	与上一卡口时间差
1	NJ100**	31000300001904	2016-01-13 11:59:45	-
2	NJ100**	31000300004302	2016-01-13 12:10:24	639 秒
3	NJ100**	31000300010504	2016-01-13 12:18:03	459 秒
4	NJ100**	31000300010504	2016-01-13 12:18:03	0 秒
5	NJ100**	31000300009004	2016-01-13 13:20:41	3758 秒
6	NJ100**	31000300011701	2016-01-13 13:33:02	741 秒
7	NJ100**	31000300019901	2016-01-13 13:35:43	161 秒
8	NJ100**	31000300020004	2016-01-13 19:21:13	20730 秒
9	NJ100**	31000300013919	2016-01-13 19:25:48	275 秒

[0066] 表2

[0067] 第3条和第4条过车记录所经过的卡口ID一致,且过车时间相同,这种属于重复记录。

[0068] 原因是由于卡口捕获车辆的时候,可能会产生多条记录。定义同一车辆,同一卡口,且时间误差不超过阈值 ΔT (本实施例设为4秒钟)的记录为重复记录,去除重复记录。

[0069] 2.2、清洗脏数据:由于号牌号码是卡口系统根据图片识别,存在一定数量的车牌字符,卡口无法识别,记录中的号牌号码不符合命名规则,具体包括:“无法识别”,“NULL”,号牌号码含有符号“?”。清洗该部分数据,部分案例如下表3所示:

[0070]

序号	号牌号码	过车时间
----	------	------

1	???????	2016-01-15 14:52:51
2	NULL	2016-01-20 19:32:30
3	宁B?711T	2016-01-25 11:31:34
4	浙A00?NT	2016-01-25 20:54:04
5	浙A025X?	2016-01-21 14:18:13
6	无法识别	2016-01-10 22:49:28

[0071] 表3

[0072] 3.3、清洗记录不齐全的卡口数据：实际情形下，由于断电、断网、卡口损坏等因素，部分卡口的数据存在缺失。对于记录不齐全的卡口数据，该卡口数据均予以清洗。

[0073] 本案例中，选取的时间区间为1月1日至1月27日，累计489个卡口中有169个卡口至少1天有数据缺失，清洗这些卡口所有的数据。

[0074] 卡口数据缺失定义如下：如果卡口一天的过车记录小于日均过车记录的1/20，该卡口这一天数据缺失。

[0075] S3. 计算卡口间流向概率

[0076] 3.1首先确定行车轨迹，并提取卡口向量

[0077] 行车轨迹确定过程：按照号牌号码，过车时间排序即可得到行车轨迹。行车轨迹包含三个部分（卡口ID，号牌号码，过车时间），表示车辆依次经过的卡口和过车时间。格式如下：本实施例共有2123568条行车轨迹，如下表4所示（省略号部分为未显示部分）。

[0078]

序号	号牌号码	卡口 ID	过车时间
1	浙 A2M1**	31000300007402	2016-01-04 07:51:09

[0079]

2	浙 A2M1**	31000300010702	2016-01-04 08:48:26
3	浙 A2M1**	31000300010904	2016-01-04 08:50:13
4	浙 A2M1**	31000300004504	2016-01-04 08:50:38
5	浙 A2M1**	31000300004502	2016-01-04 08:50:58
6	浙 A2M1**	31000300019902	2016-01-04 08:53:36
7	浙 A2M1**	31000300005402	2016-01-04 08:59:18
...

[0080] 表4

[0081] 提取卡口向量：对于同一辆车行车轨迹上相邻的两个卡口，依次取出来，作为卡口向量。对于S3案例，可取出六个向量（31000300007402, 31000300010702）、（31000300010702, 31000300010904）、（31000300010904, 31000300004504）、（31000300004504, 31000300004502）、（31000300004502, 31000300019902）、（31000300019902, 31000300005402）。

[0082] 对本实施例中的2123568条行车轨迹，分别提取向量，所有的行车轨迹可以取出的向量为85259515个。

[0083] 3.2. 计算卡口间流向概率

[0084] 统计所有的向量 (K_i, K_j) ，可以得到从卡口 K_i 流出，流向卡口 K_j 的车辆数。统计 K_i ，我们可以得到从 K_i 流出的车辆总数，得到如下表5所示(省略号部分为未显示部分)。

[0085]

卡口 K_i	卡口 K_j	count(K_i, K_j)	count(K_i)	概率
31000300000102	31000300001804	35433	156351	22.7%
31000300000102	31000300012619	35384	156351	22.6%
31000300000102	31000300001802	26530	156351	17.0%
31000300000102	31000300027001	18117	156351	11.6%
31000300000102	31000300009719	10139	156351	6.5%
31000300000102	31000300000502	5298	156351	3.4%
31000300000102	31000300000904	4236	156351	2.7%
31000300000102	31000300000503	3885	156351	2.5%
31000300000102	31000300002504	2150	156351	1.4%
31000300000102	31000300000504	1190	156351	0.8%
31000300000102	31000300000902	1180	156351	0.8%

[0086]

31000300000102	31000300025819	962	156351	0.6%
31000300000102	31000300005002	820	156351	0.5%
31000300000102	31000300012120	810	156351	0.5%
.....
.....

[0087] 表5

[0088] 卡口流向概率在另一种维度上体现了卡口分布和路网结构。如果卡口分布或者路网结构发生显著变化，该概率需要重新计算(朝夕车道、限行等不需要重新计算)。

[0089] S4. 计算行为异常车辆清单

[0090] 我们已经在S3中提取了每一辆车的行车轨迹。一个有N个点的行车轨迹对应N-1个卡口向量，每一个卡口向量，对应一个卡口流向概率，也就是说一个有N个点组成的行车轨迹对应N-1个概率值，如下表6：

[0091]

序号	号牌号码	卡口ID	过车时间	出现在该点概率
k1	浙AF1***	31000300011101	2016-01-14 07:55:58	
k2	浙AF1***	31000300005401	2016-01-14 08:06:34	25.47%
k3	浙AF1***	31000300004703	2016-01-14 08:53:38	0.00%
k4	浙AF1***	31000300010502	2016-01-14 12:12:01	0.01%
k5	浙AF1***	31000300010504	2016-01-14 12:12:34	13.07%
k6	浙AF1***	31000300007301	2016-01-14 13:59:31	0.04%
k7	浙AF1***	31000300004304	2016-01-14 16:34:15	0.28%

k8	浙AF1***	31000300004302	2016-01-14 16:34:15	53.37%
k9	浙AF1***	31000300005001	2016-01-14 16:49:02	1.71%
k10	浙AF1***	31000300014819	2016-01-14 16:57:21	0.02%
k11	浙AF1***	31000300010020	2016-01-14 17:08:16	10.55%
k12	浙AF1***	31000300028603	2016-01-14 17:17:45	16.61%
k13	浙AF1***	31000300025302	2016-01-14 17:39:38	2.19%
k14	浙AF1***	31000300004702	2016-01-14 17:45:08	0.02%

[0092] 表6

[0093] 表6中,k3、k4、k6、k10和k14概率均低于 T_i 。

[0094] 本实施例中,计算卡口阈值 T_i 的计算过程如下:

[0095] 4.1、首先统计所有卡口过车总数和号牌错误数(号牌号码不在号牌库中的车辆)。总的过车数量为:87383083,号牌错误数为:2524476;可以计算出卡口号牌识别正确率 F 为97.1%。

[0096] 4.2、对每一个卡口计算阈值 T_i ;例如对于卡口31000300015401,与其他各个卡口间的概率按照降序排列如下表7所示(省略号部分为未显示部分):

[0097]

序号	K_i	K_j	概率
1	31000300015401	31000300006701	49.594%
2	31000300015401	31000300015501	13.035%
3	31000300015401	31000300006504	11.874%
4	31000300015401	31000300006001	8.771%
5	31000300015401	31000300015601	3.733%
6	31000300015401	31000300011501	2.199%
7	31000300015401	31000300015701	1.821%
8	31000300015401	31000300006502	1.490%
9	31000300015401	31000300011601	0.957%
10	31000300015401	31000300011504	0.854%
11	31000300015401	31000300011701	0.832%
12	31000300015401	31000300011704	0.730%
13	31000300015401	31000300006904	0.728%
14	31000300015401	31000300011503	0.309%
15	31000300015401	31000300015801	0.268%
16	31000300015401	31000300006901	0.256%
17	31000300015401	31000300011703	0.255%
18	31000300015401	31000300006902	0.135%
19	31000300015401	31000300010404	0.133%
20	31000300015401	31000300011604	0.128%
21	31000300015401	31000300007003	0.096%
22	31000300015401	31000300007602	0.087%

23	31000300015401	31000300022904	0.082%
24	31000300015401	31000300010801	0.073%
• • •	• • •	• • •	• • •

[0098] 表7

[0099] 对卡口 K_i 而言, T_i 的取值应满足 $F < \sum_j P_{ij} < F + T_i$,其中 j 满足 $P_{ij} > T_i$,本案例中,如果 T_i 取值为0.268%, $\sum_j P_{ij} = 96.930\% < F$,如果 T_i 取值为0.255%, $F < \sum_j P_{ij} = 97.198\% < F + T_i$,如果 T_i 取值为0.135%, $\sum_j P_{ij} = 97.333\% > F + T_i$,所以对于卡口31000300015401,阈值 $T_i = 0.255\%$ 。

[0100] 按照如上方法,我们计算每一个卡口对应的 T_i ,得到如下表8(省略号部分为未显示部分):

[0101]

序号	卡口ID	T_i
1	31000300000704	0.262%
2	31000300000801	0.222%
3	31000300000803	0.283%
4	31000300011804	0.273%
5	31000300011919	0.236%
6	31000300011920	0.191%
7	31000300012119	0.147%
8	31000300012120	0.249%
9	31000300025301	0.146%
10	31000300020004	0.255%
• • •	• • •	• • •

[0102] 表8

[0103] 在本实施例中,我们发现该段包含14个点的行车轨迹中,有5次概率小于阈值 T_i 。

[0104] 将这14个卡口点显示到地图上,发现这14个卡口点明显分布成两条轨迹,其中 k_1 、 k_2 、 k_4 、 k_5 、 k_{10} 、 k_{11} 、 k_{12} 、 k_{13} 这8个点明显呈现一条轨迹, k_3 、 k_6 、 k_7 、 k_8 、 k_9 、 k_{14} 明显呈一条轨迹。

[0105] 嫌疑度排序: Q_i 小于阈值 T_i 一次,我们记为有一次行为异常,如果行为异常次数较多,说明该车轨迹在空间的不连续性非常高,也就是说套牌的可能性很高。重复步骤S6,我们可以得到每一辆车行为异常的次数,按照行为异常的次数降序,可以得到行为异常车辆清单,部分结果如下表9:

序号	号牌号码	总出现次数	套牌嫌疑次数
1	浙 AA59**	892	28
2	浙 A925**	475	22
3	浙 A2EM**	453	21
4	浙 A2KA**	608	21
5	浙 A255**	251	19
6	浙 AK5X**	500	18
7	浙 AC29**	160	17
8	浙 A9EN**	361	17
9	浙 A295**	556	17
10	浙 AH52**	75	16
*****	*****	*****	*****
*****	*****	*****	*****

[0108] 表9

[0109] S5. 计算车辆热点区域卡口。

[0110] 5.1、统计按照号牌分组,车辆在每一个卡口出现的次数,并且按照出现次数降序,对于车辆“浙ANOM*****”可以得到如下表10:

[0111]

序号	卡口 ID	号牌	经过次数
1	31000300020402	浙 ANOM****	46
2	31000300020403	浙 ANOM****	46
3	31000300023902	浙 ANOM****	46
4	31000300023901	浙 ANOM****	45
5	31000300011201	浙 ANOM****	35
6	31000300011202	浙 ANOM****	29
7	31000300023101	浙 ANOM****	22
8	31000300008001	浙 ANOM****	21
9	31000300004003	浙 ANOM****	18
10	31000300025201	浙 ANOM****	18
11	31000300010904	浙 ANOM****	17
12	31000300006904	浙 ANOM****	17
13	31000300015903	浙 ANOM****	16
14	31000300004403	浙 ANOM****	12
15	31000300006604	浙 ANOM****	11
16	31000300027602	浙 ANOM****	9
17	31000300027601	浙 ANOM****	9
18	31000300007402	浙 ANOM****	8
19	31000300004404	浙 ANOM****	8
20	31000300006603	浙 ANOM****	7
21	31000300010404	浙 ANOM****	6
22	31000300001904	浙 ANOM****	5
23	31000300007901	浙 ANOM****	4
24	31000300009804	浙 ANOM****	2
25	31000300023903	浙 ANOM****	2
26	31000300011101	浙 ANOM****	1
27	31000300025302	浙 ANOM****	1
28	31000300022801	浙 ANOM****	1
29	31000300011601	浙 ANOM****	1
30	31000300009803	浙 ANOM****	1

[0112]

31	31000300013820	浙 ANOM****	1
32	31000300019901	浙 ANOM****	1
33	31000300027604	浙 ANOM****	1
34	31000300009703	浙 ANOM****	1
35	31000300027701	浙 ANOM****	1
36	31000300002201	浙 ANOM****	1
37	31000300015401	浙 ANOM****	1
38	31000300007401	浙 ANOM****	1
39	31000300010503	浙 ANOM****	1
40	31000300015601	浙 ANOM****	1
41	31000300006001	浙 ANOM****	1
42	31000300005101	浙 ANOM****	1

[0113] 表10

[0114] 5.2、计算车辆对应的阈值 M_v 。对于车辆“浙ANOM****”，各个卡口点出现次数的平

均值是11.3,因此,其阈值 M_v 设置为 $11.3 \times 80\% = 9.1$ 。

[0115] 5.3、获取车辆热点活动区域卡口。根据出现次数大于均值80%这一约束,表中序号为1-15的卡口是车辆“浙AN0M****”活动热点区域卡口。

[0116] S6.计算车辆两辆热点区域卡口点之间通行的概率:

[0117] 对车辆热点活动区域的卡口点做笛卡尔积,得到卡口对,并且根据S5中的概率,给每一对卡口对进行赋值,并且选取正反向中较大的概率。

[0118] 以车辆“浙AN0M****”为例,累计15个热点活动区域卡口点,做笛卡尔积之后可以得到225个卡口对,然后给每一个卡口对赋值,例如在S5中 $P(31000300020402, 31000300023902) = 51.4\%$, $P(31000300023902, 31000300020402) = 1.4\%$,那么 $Q(31000300020402, 31000300023902) = Q(31000300023902, 31000300020402) = 51.4\%$, $Q(K_i, K_j)$ 表示卡口 K_i 和 K_j 之间无向通行概率。

[0119] S7.根据S6所得卡口间概率,对所有卡口进行聚合。

[0120] 在本实施例中,卡口通过概率连接的方式进行聚合。

[0121] Step7.1.获取可以车辆号牌清单。

[0122] Step7.2.从号牌清单中取出一个号码,根据号码,获取该号牌热点活动区域卡口($K_1, K_2, K_3 \dots K_n$),记为数组A0,并且根据车辆出现在该卡口的次数倒叙排列。

[0123] Step7.3.将A0中的第一个卡口 K_1 添加到数组A1中,并且将 K_1 从A0中去除。

[0124] Step7.4.判断 K_1 与A0中每一个卡口之间的无向概率 $Q(K_1, K_i)$ 是否大于阈值 T_i ,如果是,将 K_i 添加到A1和空数组A_new,并且将 K_i 从数组A0中去除。

[0125] Step7.5.从数组A_new中取出一个卡口 K_j ,判断 K_j 与A0中每一个卡口之间的无向概率 $Q(K_j, K_i)$ 是否大于阈值 T_i ,如果是,将 K_i 添加到A1和空数组A_new,并且将 K_i 从数组A0中去除,将 K_j 从数组A_new中去掉。

[0126] Step7.6.重复Step7.5,直到A_new为空。

[0127] Step7.7.增加数组A_i,重复步骤Step7.3-7.6。

[0128] Step7.8.重复Step7.7,直到A0为空。

[0129] Step7.9.重复Step7.2-7.8,直到号牌清单为空。

[0130] 通过以上运算,可以获取所有车辆热点活动区域卡口点的聚合情况。

[0131] S8.根据卡口点的聚合情况,计算车辆套牌概率Z。

[0132] 如果所有卡口都分到A1组,也就是说车辆所有热点活动区域卡口都能够连接起来,如图3所示,那么说明车辆经常活动的区域是处在一个连接程度比较紧密的地方,该车辆套牌可能性较小。

[0133] 如果卡口比较均匀的分到A1组合A2组中,说明车辆热点活动区域能够明显的分成两个社区,如图4所示,那么说明车辆活动在两个连接程度较低的区域,同时考虑到该车辆有多次不正常的跳转行为,说明该车辆套牌可能性较高。

[0134] 如果卡口分布在多个组中,说明可能有多个车牌误识别成该辆车,分组越多,说明该车辆的套牌可能性越小。

[0135] 在本实施例中,部分计算结果如下表11(按照行为异常次数排序):

[0136]

序号	号牌号码	行为异常次数	分组 A1	分组 A2	分组 A3	分组 A4	I	S	Z
1	浙 A6C***	400	25	0	0	0	-	1.00	-
2	浙 A6F***	182	15	0	0	0	-	1.00	-
3	浙 A8B***	158	14	0	0	0	-	1.00	-
4	浙 A6C***	139	13	0	0	0	-	1.00	-
5	浙 A11***	134	22	3	1	0	0.50	0.14	0.00
6	浙 A6C***	134	2	0	0	0	-	1.00	-
7	浙 A6C***	130	6	5	5	2	-	0.83	-
8	浙 A0C***	128	11	0	0	0	-	1.00	-
9	浙 A1B***	123	6	0	0	0	-	1.00	-
10	浙 A2A***	117	2	0	0	0	-	1.00	-
11	浙 A3B***	113	1	2	1	6	-	0.33	-
12	浙 A2A***	111	4	2	5	0	0.50	0.80	2.25
13	浙 A62***	110	12	0	0	0	-	1.00	-
14	浙 A6F***	109	2	0	0	0	-	1.00	-
15	浙 A4G***	108	2	7	1	1	-	0.29	-
16	浙 A2A***	106	4	2	4	0	0.50	1.00	5.00
17	浙 A1B***	106	48	0	0	0	-	1.00	-
18	浙 A6C***	104	13	0	0	0	-	1.00	-
19	浙 A18***	104	6	0	0	0	-	1.00	-
20	浙 A19***	103	46	0	0	0	-	1.00	-
21	浙 A2L***	103	5	0	0	0	-	1.00	-
22	浙 A6C***	98	13	0	0	0	-	1.00	-
23	浙 A10***	98	55	0	0	0	-	1.00	-
24	浙 A1A***	96	71	0	0	0	-	1.00	-
25	浙 A6C***	96	12	0	0	0	-	1.00	-
26	浙 A6C***	95	12	0	0	0	-	1.00	-
27	浙 A3B***	94	1	0	0	0	-	1.00	-
28	浙 A0A***	93	3	4	0	0	1.00	0.75	2.21
29	浙 A2A***	93	5	0	0	0	-	1.00	-
30	浙 A18***	92	35	0	0	0	-	1.00	-
31	浙 A6C***	91	12	0	0	0	-	1.00	-
32	浙 A1B***	91	48	0	0	0	-	1.00	-
33	浙 A3F***	91	14	0	0	0	-	1.00	-
34	浙 A18***	89	43	2	1	0	0.50	0.05	0.00
35	浙 A96***	89	9	0	0	0	-	1.00	-
36	浙 A14***	89	51	2	0	0	1.00	0.04	0.00
37	浙 A2A***	86	4	2	1	1	-	0.50	-
38	浙 AE2***	86	38	3	0	0	1.00	0.08	0.00
39	浙 A6C***	86	12	0	0	0	-	1.00	-
40	浙 A80***	85	1	2	0	0	1.00	0.50	0.19
41	浙 A6C***	84	11	0	0	0	-	1.00	-
42	浙 AE2***	84	68	0	0	0	-	1.00	-
43	浙 A96***	84	8	0	0	0	-	1.00	-
44	浙 AW0***	83	41	6	0	0	1.00	0.15	0.02
45	浙 A6C***	81	3	2	4	1	-	0.75	-
46	浙 A14***	81	75	0	0	0	-	1.00	-

[0137]

47	浙 A6C***	81	15	0	0	0	-	1.00	-
48	浙 A8G***	81	9	0	0	0	-	1.00	-
49	浙 A0F***	80	13	0	0	0	-	1.00	-
50	浙 A10***	80	52	2	0	0	1.00	0.04	0.00
51	浙 A10***	80	12	0	0	0	-	1.00	-
52	浙 A0F***	79	14	0	0	0	-	1.00	-
53	浙 A03***	78	26	2	1	0	0.50	0.08	0.00
54	浙 A96***	76	7	0	0	0	-	1.00	-
55	浙 A96***	76	8	0	0	0	-	1.00	-
56	浙 A74***	76	53	0	0	0	-	1.00	-
57	浙 A98***	74	1	2	1	3	-	0.67	-
58	浙 A18***	74	44	2	0	0	1.00	0.05	0.00
59	浙 A2L***	74	18	0	0	0	-	1.00	-
60	浙 A0F***	74	11	0	0	0	-	1.00	-
61	浙 A6C***	74	5	3	1	1	-	0.60	-
62	浙 AZD***	73	51	0	0	0	-	1.00	-
63	浙 A0A***	73	8	0	0	0	-	1.00	-
64	浙 A62***	73	15	0	0	0	-	1.00	-
65	浙 A12***	72	8	0	0	0	-	1.00	-
66	浙 A62***	72	14	0	0	0	-	1.00	-
67	浙 A3B***	72	15	0	0	0	-	1.00	-
68	浙 A0F***	71	4	5	0	0	1.00	0.80	3.69
69	浙 A0C***	71	4	0	0	0	-	1.00	-
70	浙 A3B***	71	3	3	0	0	1.00	1.00	6.00
71	浙 A10***	71	49	2	0	0	1.00	0.04	0.00
72	浙 A6C***	70	13	0	0	0	-	1.00	-
73	浙 A62***	70	7	2	1	1	-	0.29	-
74	浙 A6F***	70	16	0	0	0	-	1.00	-
75	浙 A96***	70	7	0	0	0	-	1.00	-
76	浙 A98***	69	4	16	0	0	1.00	0.25	0.08
77	浙 A6C***	69	15	0	0	0	-	1.00	-
78	浙 AF1***	69	12	13	1	0	0.50	0.92	9.44
79	浙 A2A***	69	4	1	0	0	1.00	0.25	0.02
80	浙 A62***	69	11	0	0	0	-	1.00	-
81	浙 A2L***	69	15	0	0	0	-	1.00	-
82	浙 A7B***	68	41	0	0	0	-	1.00	-
83	浙 A05***	68	6	2	5	2	-	0.83	-
84	浙 A2L***	68	5	0	0	0	-	1.00	-
85	浙 AF8***	68	43	2	1	0	0.50	0.05	0.00
86	浙 A7A***	68	4	45	1	0	0.50	0.09	0.00
87	浙 A10***	68	68	0	0	0	-	1.00	-
88	浙 A3F***	67	14	0	0	0	-	1.00	-
89	浙 A6C***	67	11	0	0	0	-	1.00	-
90	浙 A6C***	67	14	0	0	0	-	1.00	-
91	浙 A85***	67	28	0	0	0	-	1.00	-
92	浙 A3B***	67	8	0	0	0	-	1.00	-
93	浙 A1A***	66	55	0	0	0	-	1.00	-
94	浙 A1U***	66	38	0	0	0	-	1.00	-
95	浙 A18***	66	41	2	0	0	1.00	0.05	0.00

[0138]	96	浙 A81***	66	2	8	1	1	-	0.25	-
	97	浙 A13***	65	86	0	0	0	-	1.00	-
	98	浙 A13***	65	85	0	0	0	-	1.00	-
	99	浙 A55***	65	6	0	0	0	-	1.00	-
	100	浙 A0F***	65	11	0	0	0	-	1.00	-

[0139] 表11

[0140] 本实施中,当嫌疑度 Z 大于等于1的时候,我们认为该车具有较高的套牌概率,我们可以发现,在行为异常次数最多的前100辆车当中,仅有6辆车嫌疑度 Z 大于1,也就是把100辆车的排查范围缩小到了6辆车。

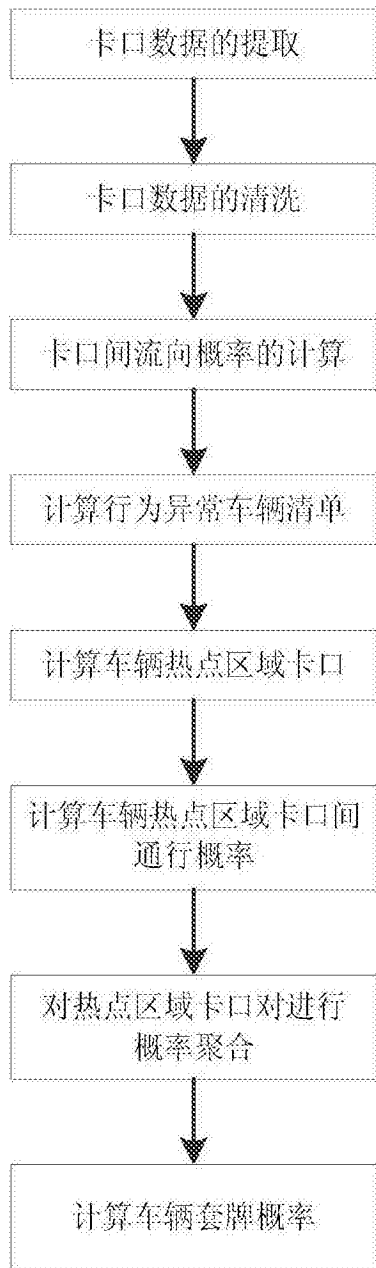


图1

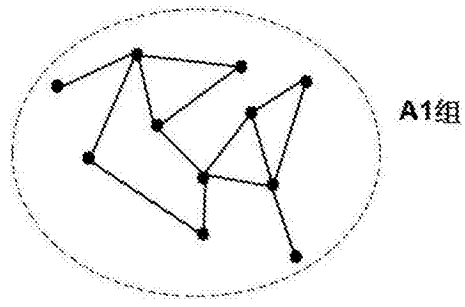


图2

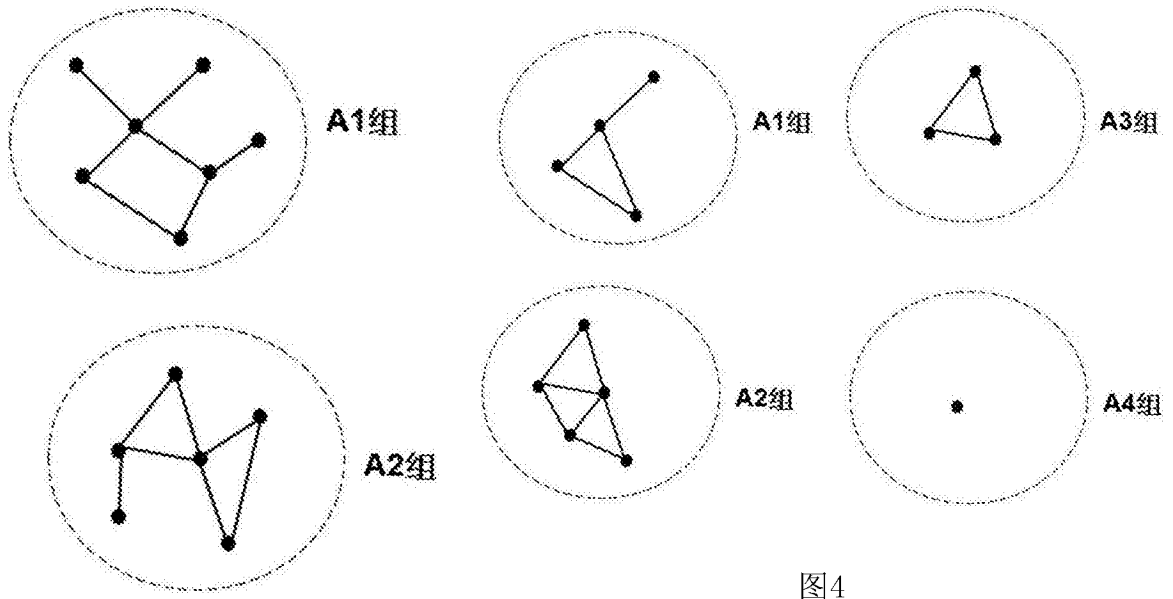


图3

图4