DOI: 10.11992/tis.201905058
网络出版地址: http://kns.cnki.net/kcms/detail/23.1538.tp.20190909.1300.004.html

基于区块链的公共数据电子证据系统及关联性分析

李萌[1,2], 刘文奇[1,2], 米允龙[3]
（1. 昆明理工大学 理学院, 云南 昆明 650500; 2. 昆明理工大学 数据科学研究中心, 云南 昆明 650500; 3. 中国科学院 大数据挖掘与知识管理重点实验室, 北京 100190）

摘　要: 针对公共部门提供电子证据时, 必须保证数据的真实性和证明力, 同时要尽量保护当事人隐私和他人利益, 而从公共数据库中提取电子数据并形成有效电子证据, 既是法律难题也是技术难题这一问题, 本文在电子证据可信性的影响因素分析基础上, 提出了自动生成中国公共数据库电子证据系统的区块链模型的取证技术体系, 并从司法角度出发, 提出了公共数据治理的电子证据生命周期、内容关联、载体关联和智能串并分析方法。本文的研究在一定程度上保证了司法、公证和公共事务中电子证据的可信性, 并实现了从公共数据库中自动生成证据知识的原型系统。

关键词: 公共数据库; 电子证据; 区块链; 关联分析; 智能串并
中图分类号: TP391　　**文献标志码:** A　　**文章编号:** 1673−4785(2019)06−1127−11

中文引用格式: 李萌, 刘文奇, 米允龙. 基于区块链的公共数据电子证据系统及关联性分析 [J]. 智能系统学报, 2019, 14(6): 1127–1137.
英文引用格式: LI Meng, LIU Wenqi, MI Yunlong. An electronic evidence system based on blockchain and correlation analysis[J]. CAAI transactions on intelligent systems, 2019, 14(6): 1127–1137.

An electronic evidence system based on blockchain and correlation analysis

LI Meng[1,2], LIU Wenqi[1,2], MI Yunlong[3]
(1. Faculty of Science, Kunming University of Science and Technology, Kunming 650500, China; 2. Center of Data Science, Kunming University of Science and Technology, Kunming 650500, China; 3. Key Laboratory of Big Data Mining and Knowledge Management, Chinese Academy of Sciences, Beijing 100190, China)

Abstract: When the public sector provides electronic evidence, the authenticity and proof of the data must be ensured, and meanwhile, the privacy of the concerned parties and the interests of others should be protected as much as possible. Therefore, forming effective electronic evidence by extracting electronic data from public databases is not only a legal problem but also a technical problem. Based on the analysis of the influencing factors of the credibility of the electronic evidence, the automatic blockchain model of an electronic data system based on the Chinese public database is proposed in this paper. In addition, the methods of life-cycle management, content association, carrier association, and intelligent string-concatenation for electronic evidence based on public data are put forward from the legal viewpoint. To a certain extent, this research guarantees the credibility of electronic evidence in judicial, notary, and public affairs, and implements a prototype system for automatically generating evidence knowledge from public databases.

Keywords: public database; electronic evidence; blockchain; correlation analysis; intelligent serial and parallel

中国公共数据库系统发展到一定阶段, 公共数据资源将成为国家战略资产。有效的数据治理是数据资产形成和高效利用的必要条件[1]。公共数据治理是指在公共部门、公民、企业和社会组织所提供的零散数据基础上形成统一的、可信的主数据, 并且依法有序地提供给公共部门、公众、企业和社会组织综合运用的过程。

为公众提供优质的公共数据库数据服务是公共数据当局的首要任务, 提供高可信度的电子数据是最重要的任务。其中有一大类电子数据将用作相应的公共产品的证据要件, 如居民户籍证明、婚姻证明等, 乃至司法和仲裁等维护社会公

收稿日期: 2019−05−28.　**网络出版日期:** 2019−09−10.
基金项目: 国家自然科学基金项目 (61573173).
通信作者: 刘文奇. E-mail: liuwenq2215@sina.com.

平正义的一类公共产品中的电子证据。与一般的商业数据治理相比，公共数据治理更多地涉及法律适定性问题。公共数据库数据的电子证据功能涉及很多方面，比如电子证据的生命周期管理以及电子证据存储和归档格式、电子证据与案件事实的关联性等。这类具有电子证据性质的公共数据产品的基本要求是真实可信性及其与现实案件的高度关联性。因此，构造多层的、分布式且防篡改算法和安全的电子证据取证系统是基于公共数据库的电子证据的关键技术。以求解拜占庭将军问题的算法为基础发展起来的区块链技术在去中心化安全技术领域已经取得一定成效，如比特币区块链技术、能源互联网区块链技术等[2-3]。此外，在医疗保健中广泛地实施区块链以提高数据隐私性、互操作性和可扩展性[4-5]。在司法和公共安全信息技术领域，分布式的电子证据广泛存在，急需建立更加严密的电子证据信任技术体系。

按物证说的观点，相对于传统物证而言，电子证据产生和存在的方式有很大的区别，主要体现在电子证据的符号化、易篡改性、可删除性、可分离性、易复制性、易破坏性，使得电子证据在收集和使用过程中真实性会发生改变。狭义的电子证据是指以存储于介质载体中的电磁记录或光电记录并对司法案件审理、仲裁等事实起证明作用的电子数据 (含视听资料) 及其附属物。除了具有证据的客观性和可知性之外，电子证据还具有非直观性和多态性、电子物理和诉讼证据的多重属性。为了保持电子证据的客观真实性，在获取电子证据时，应采用取证专用的数据拷贝机和电子证据勘验取证技术，附加上时间戳数据，一次性提取和固定介质载体中的全部电子数据。广义的电子证据是指，用于公共管理、认证认可、司法、仲裁、公证等事务的电子数据及其附属物。广义的电子证据与狭义的电子证据相比，应用范围更宽、取证过程相对简单。

电子证据与传统证据相比，最突出的特点是：1) 它需要借助一定的介质存储，通常存储于电子设备的存储器中；2) 电子证据不能直观查看，必须借助适当的电子系统软硬件环境显示后才能查看；3) 由于电子证据存在于虚拟空间，所以可以迅速传播并且精确复制；4) 电子证据中的数字证据很容易被修改或删除，并且不易找到更改“痕迹”；5) 有些电子证据有时限性，可能随时间而消失[6]。

电子数据的真实性、合法性、关联性和证明力是电子证据的 4 个维度。根据刘品新[7] 的研究，在司法实践中电子证据被质疑的几率是比较高的。在司法实务中，电子证据的攻防成效令人堪忧，司法运用电子证据尚未形成成熟的机制。鉴于电子证据易受质疑，中华人民共和国最高检察院对电子数据和视听资料的审查、认定和是否作为定案依据有明确的规定，严格要求对电子数据和视听资料结合案件的其它证据审查其真实性和关联性。

由此可见，电子证据的关联性和真实性是电子证据在法庭上是否被采信的关键性指标。作为一种虚拟空间的证据，电子证据用于定案必须同时满足内容上以及载体上的关联性和真实性。互联网、物联网与大数据的出现和发展在极大提高公众和公共部门数据交互效率的同时，也为保障电子证据的真实性、关联性带来了新的挑战。由于公共部门以维护社会公平正义为目的，因此公共数据库的电子证据从内容到载体都应该是电子证据中最为可信的。

传统的公共数据库中数据的应用需求主要包含公共产品供给和消费过程的记录。但是随着网络和智能终端的日益普及，公共数据的边界日益扩大，大量的公共数据的电子化，纳入了海量的机器数据，这将带来公共数据库中的电子数据证据功能复杂化。随着公共数据库数据边界的扩张，公共数据库的电子证据功能将成为公共数据库主要功能之一。如公共安全数据库中的涉案物品记录、消防数据、环境监测、宾馆饭店住宿记录、出租车定位记录以及医疗健康数据库中的电子病历、防疫检疫记录等。

公共数据库的电子证据系统的应用与法律密切相关，可信性是必然要求。在中国的法律框架之下，数据必须满足：1) 及时性，数据必须是及时收集的；2) 过程性，过程的数据必须被记录；3) 不可篡改性，所收集及存储的数据必须证明没有被篡改过。

其中不可篡改性是电子证据的特性，也是电子证据系统设计的关键技术难点。不可篡改性有两个环节：1) 公共数据库内部的电子证据生成过程的不可篡改性，即电子证据的保障品质，或保质；2) 电子证据的外部转移与再现过程的不可篡改性，即电子证据的保障安全，或保全。

在中国的法律中，电子数据、电子证据概念经常混合使用。在司法、仲裁和行政案件处理实务中，虽然可以作为证据使用 (电子证据、电子书证或视听电子材料)，但是单一的电子证据并不能

够作为判定事实的根据, 电子证据需要跟其他证据一起使用, 并且可以相互印证, 因此组成证据链条来证明案件事实。电子证据有效的前提是电子数据本身是可信的。区块链技术在解决共享经济中的信任问题方面颇具潜力, 在某种程度上区块链适合取代第三方的信任[8-9]。

公共数据库是公共产品, 公共部门是其供给者。公共部门有义务依法从其主导的公共数据库中提供公民和法人所需要的一切电子证据 (证明)。当公民和组织必须需要公共数据库中的电子数据维护自身权益的时候, 公共部门必须向他们提供具有完整法律效力的电子证据。这些电子证据的运用可能不仅限于司法事务。在公共部门提供电子证据时, 必须保证数据的真实性和证明力, 同时要尽量保护当事人隐私和他人利益。因此, 从公共数据库中提取电子数据并形成有效电子证据既是法律难题也是技术难题。

本文将构建自动生成中国公共数据库电子证据系统的区块链模型及其取证技术体系, 在一定程度上保证了司法、公证和公共事务中电子证据的可信性。最后提出了公共数据治理的电子证据生命周期管理并对电子证据的关联性进行了分析。

1　电子证据可信性的影响因素分析

电子证据的不可篡改性包括数据的保质和保全, 它与传统证据的有效性与证据保全相对应, 具体体现在电子证据的数据攫取、固定、保管、转移等各个环节。但与传统证据相比, 电子证据的产生和存在的形式完全不同[10]。由于电子数据科技含量高、易篡改、可分离等特点, 使之非常容易被修改、伪造和删除, 加大了电子证据的保质和保全难度, 仅仅通过法律措施和公证机关很难有效控制电子证据的法律效力。从普通证据学的原理来说, 司法实践中对传统证据认定普遍采用正面认定法和侧面认定法, 其中正面认定法是主要方法。参照传统证据的认定, 电子证据的正面认定须保证电子数据的可靠性, 在其运行的各个环节都有辅助证据 (如数据标签、时间戳) 加以证明, 形成电子数据保管锁链。由于电子数据通常是潜在的且与大量的无关信息纠缠共存, 有时甚至已经被删除, 故需要通过专门的技术手段发现和获得有价值的证据信息, 这就是电子证据检验技术。电子证据检验技术是对电子设备中存在的电子数据 (电子证据) 进行识别、发现、提取、保存、恢复、展示、分析和鉴定的一种全过程的科学技术, 其检验结果可作为法庭证据或案件侦查线索。

电子证据正面认定的审查需要以下环节:

1) 生成环节, 即电子证据中的数据是如何生成的。这一环节审查电子证据数据是设备采集还是人工录入。如果是设备采集的则进一步确认是由人工使用设备采集还是设备自动采集。如果是人工使用设备采集, 则需要确认采集者是否具备采集资格和设备是否正常。如果是机器自动采集数据, 则需认定设备是否正常。采集人员和设备是否正常, 则需要合法的第三方认定或检测机构相关电子文书。

2) 获取的方式。审查内容包括: 采集过程是否合法, 采集方法是否科学、可靠, 采集过程是否得到被采集方认可。

3) 传输环节。审查电子证据的数据形成过程和传输过程中使用的计算机网络或专用设备是否正常, 传输过程中电子证据的数据是否可能被修改, 传输过程中数据是否被非法复制、截取。

4) 存储环节。该环节审查电子证据数据是怎样存储的, 是否科学, 存储介质与存储过程是否安全可靠, 是否以加密形式存储, 存储后是否有访问权限上的漏洞, 存储中是否有非法篡改和销毁的风险。

2　电子证据系统的区块链数据模型

关于电子证据系统的保质问题, 我国迄今为止没有法律规定, 也没有完整的行政规范。为了解决公共数据库中可能用于电子证据的数据的可信性, 必须建立公共数据库的全局信任机制。有效的解决办法是在公共数据库中建立区块链系统的“智能合约”层, 即建立一种无法被篡改和操控的“代码合同”[11]。智能合约并非法律所界定的合同, 而是执行在区块链上的代码, 故也称作“链上代码”。为了实现中国公共数据库中用于提供电子证据的部分数据的法律效力, 这种链上代码必须遵从不可篡改性和法律上的可验证性。电子证据系统在公共数据库数据生产过程中提取的数据在数据博弈参与者之间形成区块链, 其分布式账本将保证数据的一致性、不可篡改性和合法性[12]。在事务方式上, 电子证据系统的区块链的每个节点上都有自己的本地数据库。

根据电子证据系统的上述要求, 我们构建一种基于区块链的数据安全共享网络体系, 如图 1 所示。

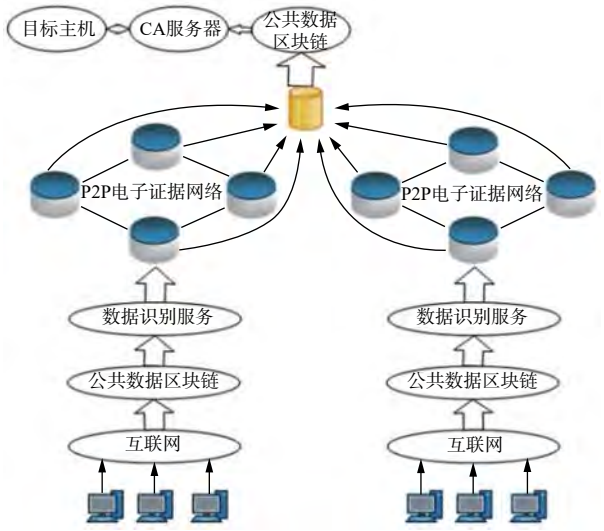

图 1　电子证据区块链结构

Fig. 1　Blockchain structure of the E-evidence system

该体系依赖于现有的数据仓库架构[13-14]，将电子数据作为一种资产进行统一标识，利用区块链技术将数据进行分布式存储，通过设计高效分发协议，实现数据在参与者之间的自主对等的 P2P 电子证据网络 (peer to peer i-evidence network, P2PIEN)。该电子证据网络依托于公共数据库网络的物理系统和数据博弈覆盖网络，在逻辑上遵从电子证据系统的法律要求，并且将部分公共数据库系统事务流程去中心化[15]。本质上讲，电子证据系统是将法规所要求部分公共电子数据本地备份并形成共识节点。P2PIEN 的具体内容参见文献 [16]。

例 1　学历认证的电子证据系统原型的核心区块链。学历认证的电子证据区块链系统原型主要由从教育公共数据库 (目前包括校级、省级及国家级数据库) 中自动获取数据、形成区块并将数据加载进区块链中及进行学历认证过程 3 部分组成。

1) 获取数据阶段：将学生信息按学号自动从数据库提取出来，提取出的同一学生的不同粒度的记录将会起到相互佐证作用。获取公共数据库记录部分源码如图 2。

```
class Student:
    def _init_(self):
        self._helper=MysqlHelper();
    def getOne(self,stuNo); #外面传入参数
        sql= "select*from t_school where stuNo=%s";
        params=(stuNo,);
        return self._helper.getOne(sql, params);
    def getAll(self); #获取所有学生数据
        sql= "select stuNo from t_school";
        return self. Helper.getAll(sql);
    def checkValidate(self, stuName,stuNo):
        sql="select * from t_school where stuName=%s and stuNo=%s";
        params=(stuName, stuNo);
        return self._helper.getOne(sql, params);
```

图 2　获取数据部分源码

Fig. 2　Get the data part of the source code

2) 数据加入区块链阶段：首先，将提取的每一条学生信息形成一个区块；其次，各用户对该区块进行共识；最后，一旦共识成功，对区块进行链接，形成完整的区块链。通过共识算法来保证用户账本之间的一致性，即分布式记账核心。将具体的区块加入区块链过程如图 3~6。

```
def main();
    book=TransactionBook();
    #1.从学校将学生数据从数据库加载进区块链中
    print ("1.开始从学校信息将学生数据加载进区块链内！"):
    student=Student();
    allReconds=student.getAll();
    for stuNo in allReconds;
        if stuNo:
            data=student.getAll.getOne(stuNo);
            b=Transaction("来自学校系统", data);
            book.addBlock(b);
            print(book);
            print("学号：{}的同学已经加载成功！".format(stuNo[0]));
        else:
            print("此学生不存在！");
```

图 3　数据加入区块链

Fig. 3　Data is added to the blockchain

```
1.联盟区块链构建。
机构1:校级
已经收到信息!
加载区块信息，请稍后!
校验成功!
Hash：0000297e77f1fd64a41632b5469ca386070399aacdaf0657641b8413a2585605
StudentChains<1 Blocks, Head: 5a7a15c3bd194532bbc0cd66336el59b>
学号：200818008629001的同学已经加载成功!
已经收到信息!
加载区块信息，请稍后!
校验成功!
Hash：0000685fa261abcc29alc240b5a093d56f501a5402eab5dba9aaff71674daf17
StudentChains<2 Blocks, Head：25114c07712d45fbbe975c3c6e271c47>
学号：200818008629002的同学已经加载成功!
已经收到信息!
加载区块信息，请稍后!
校验成功!
```

图 4　校级联盟区块链

Fig. 4　School-level alliance blockchain

```
机构2:省级
已经收到信息!
加载区块信息，请稍后!
校验成功!
Hash：0000c218a43057ea3ale3a4c5ccc08d3cf43a318f3c4b64el6658347d9560494
StudentChains<5 Blocks, Head：234c254fd67d4aef944c6aea05b49214>
学号：200818008629001的同学已经加载成功!
已经收到信息!
加载区块信息，请稍后!
校验成功!
Hash：000016e2b0ca5472d5899cl2717fl7b0ba30e0b9220bbca416cl9887fed909aa
StudentChains<6 Blocks, Head: b8057635628d4f44b2fbaae44e2288d3>
学号：200818008629002的同学已经加载成功!
已经收到信息!
加载区块信息，请稍后!
校验成功!
```

图 5　省级联盟区块链

Fig. 5　Provincal-level alliance blockchain

```
机构3:国家级
已经收到信息!
加载区块信息，请稍后!
校验成功!
Hash：000037a8a54ce82d4648fl2a43e4c2dc8919f888b2el64c35870b83362f3d0d9
StudentChains<11 Blocks, Head: 30522e7fc7294fd0b3ddl8df0cc9a315>
学号：200818008629001的同学已经加载成功!
已经收到信息!
加载区块信息，请稍后!
校验成功!
Hash：00009c3e2c01c378ffl3e5a4567532c9d0a61bal2af49b520a46faac31431c6f
StudentChains<12 Blocks, Head: 202783f9ebd748a7bb761a1718362790>
学号：200818008629002的同学已经加载成功!
已经收到信息!
加载区块信息，请稍后!
校验成功!
```

图 6　国家级联盟区块链

Fig. 6　National-level alliance blockchain

3) 进行学历认证过程: 一方面, 尽管任何一级的数据库中的数据可以被管理员进行修改/删除, 但是修改后的数据记录只能以新的区块形式加入对应的链条中去; 另一方面, 当试图修改/删除已形成的区块的数据, 将导致其他用户账本对应的区块信息不一致, 这是不允许发生的。图 7~8 显示的是区块链中的数据与某账本改变区块数据失败验证。

```
2.机构读取区块链中的信息。
时间: 2006-11-09 07:14:47;来自学校系统:学生信息: (1, '200818008629001', '李小路', 0, datetime.date(2008, 9, 11),
时间: 2006-11-09 07:14:48;来自学校系统:学生信息: (2, '200818008629002', '张周', 1, datetime.date(2008, 9, 11),
时间: 2006-11-09 07:14:49;来自学校系统:学生信息: (3, '200818008629003', '萧然', 0, datetime.date(2008, 9, 11),
时间: 2006-11-09 07:14:51;来自学校系统:学生信息: (4, '200818008629030', '白易', 1, datetime.date(2008, 9, 11),
时间: 2006-11-09 07:14:51;来自省级系统:学生信息: (1, '200818008629001', '李小路', datetime.date(2008, 9, 11), '1
时间: 2006-11-09 07:14:51;来自省级系统:学生信息: (2, '200818008629002', '张周', datetime.date(2008, 9, 11), '10
时间: 2006-11-09 07:14:53;来自省级系统:学生信息: (3, '200818008629003', '萧然', datetime.date(2008, 9, 11), '10
时间: 2006-11-09 07:14:53;来自省级系统:学生信息: (4, '200818008629030', '白易', datetime.date(2008, 9, 11), '10
时间: 2006-11-09 07:14:54;来自省级系统:学生信息: (5, '200917008627020', '张丽', datetime.date(2009, 9, 11), '10
时间: 2006-11-09 07:14:54;来自省级系统:学生信息: (6, '201717008627020', '王成', datetime.date(2017, 9, 11), '10
时间: 2006-11-09 07:14:55;来自国家级信息系统:学生信息: (1, '200818008629001', '李小路', '云南省', '10674', '毕业'
时间: 2006-11-09 07:14:55;来自国家级信息系统:学生信息: (2, '200818008629002', '张周', '云南省', '10674', '毕业')
时间: 2006-11-09 07:14:56;来自国家级信息系统:学生信息: (3, '200818008629003', '萧然', '云南省', '10674', '毕业')
时间: 2006-11-09 07:14:56;来自国家级信息系统:学生信息: (4, '200818008629030', '白易', '云南省', '10674', '退学')
时间: 2006-11-09 07:14:57;来自国家级信息系统:学生信息: (5, '200917008627020', '张丽', '云南省', '10673', '毕业')
时间: 2006-11-09 07:14:57;来自国家级信息系统:学生信息: (6, '201717008627020', '王成', '云南省', '10673', '在校')
```

图 7　读取区块链中的信息

Fig. 7　Read the information in the blockchain

```
3.机构欲修改区块链中信息。
原信息为: ('200818008629030', '白易', 1, datetime.date(2008, 9, 11), '计算机应用技术', '昆明理工大学', '退学')
欲修改信息为: ('200818008629030', '白易', 1, datetime.date(2008, 9, 11), '计算机应用技术', '昆明理工大学', '毕业')
正在校验!
修改信息失败!
```

图 8　修改信息失败

Fig. 8　Modifying information failed

3　电子证据取证系统

3.1　电子证据取证系统的构成

电子证据包括取证任务生成、物理介质、取证认证、电子数据和电子证据提交。电子证据本质上仍然是计算机产生的数据, 在传输和存储过程中表现为 0 和 1 构成的字符串。在电子证据取证和保全过程中, 需要设置 CA 服务器。通过运用信息安全技术生成对电子证据本身具有证明作用的辅助证据, 形成电子证据的链锁。电子证据采集和保管系统由一个 C/S 架构的软件系统和相应的硬件部署, 外加便携式 U 盘取证工具组成。硬件部署包括 CA 服务器、数据库服务器、工作主机、U 盘取证终端。

1) CA 服务器: 提供对系统用户 (如法院、检察院等) 的注册和认证, 项目和任务的认证和授权, 任务证书的生成、签发, 电子证据加密密钥和签名密钥的生成和发放等服务。

2) 数据库服务器: 提供对系统用户数据库、项目任务数据库和电子证据数据库的服务支持。其中电子证据数据库是公共数据库中区块链的本地数据服务器, 负责相应各级数据提供者的数据变更和本地备份, 将受到智能合约的控制。

3) 工作主机: 将安装系统客户端软件, 供用户登录系统, 也是 U 盘取证工具端与 CA 服务器之间交互的桥梁, 为二者提供通信和数据传输服务。

4) U 盘取证工具端: 工具端是安装 WinPE 操作系统的导引 U 盘, 内置的 WinPE 操作系统镜像集成了为提取目标主机的计算机取证信息的数据采集软件, 同时以外部数据形式放置取证任务证书。通过 U 盘取证终端提取到的电子证据经过签名和加密处理, 以 U 盘为载体将数据转移至目标主机。

WinPE 系统是一个组件精简版的操作系统内核镜像, 其工作原理是将镜像加载至内存后以解压的方式将操作系统安装在内存中, 而不用启动主机自身的操作系统。目标主机的硬盘对于 WinPE 来说就是一块完整的外部硬盘。因此, WinPE 系统启动过程中并不使用主机的任何硬盘数据。从而保持了目标主机硬盘的完整性, 避免了一些高科技犯罪行为利用程序设置非本人进入的使用销毁程序毁灭关键证据。同时, 目标主机中的病毒、木马无法影响到取证工具端中的 WinPE 系统和文件, 从而在一定程度上保证了整个电子证据取证系统的安全性。公共数据库电子证据取证流程如图 9。

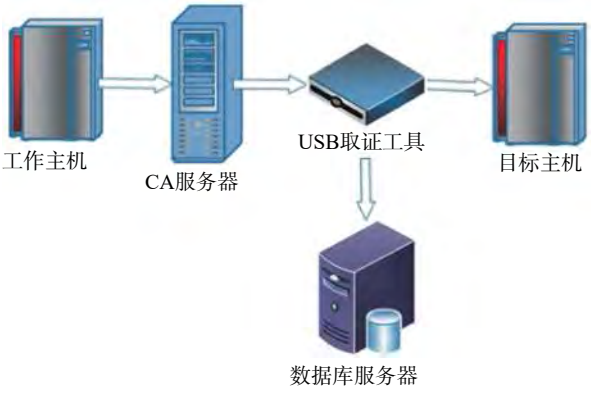

图 9　公共数据库电子证据取证流程图

Fig. 9　Forensics process of the E-evidence network

3.2　电子证据数据的生命周期管理

随着公共数据库数据规模的急速膨胀，公共部门数据资产的持有成本也迅速增长，导致公共财政中信息技术和管理预算压力持续增加。一方面，与其他公共产品供给一样可以通过适度市场化来缓解财政预算压力。另一方面，过度的收集和囤积数据造成的财政资源浪费则需要制定合理的公共数据生命周期政策，从而降低法律风险和信息技术成本。公共数据生命周期政策的基本目标包括：

1) 基于法律和制度要求，明确公共数据保留时间表。国家必须出台有关数据保留时间的法规。例如，电信数据包含有关人际关系和位置的丰富知识，对执法和反恐等公共产品具有极大作用，很多国家和地区对电信用户姓名、出生日期、计费地址、绑定银行卡号、电话号码、国际移动设备识别码、主叫号码、被叫号码、通话时长、地理位置、登录时间、下线时间、IP 地址、发件地址、访问的 URLs 等有明确的电信保留时间。中国公共数据库的数据保留时间并无明确的规定，更无完善的法律和制度。

2) 提供法律和制度保留区，并支持电子数据显示。随着公民意识的增强，大多数公共部门和非政府组织将面临要求保存证据的起诉和调查。在公共数据周期管理中心，为了应对证据收集和分析，公共数据治理计划必须控制法律风险，提供必要的电子证据展示。例如，2015 年 8 月 12 日，位于天津市天津港的瑞海公司危险品仓库发生火灾爆炸事故，造成 165 人遇难。该事故的调查报告显示，通过调取分析位于瑞海公司北侧的环发通讯公司监控视频、提取比对现场痕迹证据、分析集装箱毁坏和位移特征，认定事故最初起火部位为瑞海公司危险品仓库运抵区南侧集装箱区的中部。这个案件调查过程中，需要公安部门提交视频监控设备和消防设备的传感数据的法律效力展示。

3) 压缩公共数据并将其存档，降低信息技术成本和提高公共数据资源应用绩效。公共数据当局需要压缩静态公共数据，并将其归档，以降低存储成本，提高应用绩效。这些静态数据可能存在于数据仓库的某个环节，或分布式数据库的某个节点，包括文件系统、NoSQL 数据库，甚至包括 Hadoop 中的智能电表读数、传感器数据、RFID 数据和网络日志。在数据周期管理中，公共数据当局必须考虑数据归档适用于所有属地法规。例如，税务部门的数据当局在进行数据压缩和存档过程中，必须以电子表单生成在存档过程中保留原有格式，而不是将结构化数据转为 PDF 格式，其目的是让涉税案件中方便税务审计员确定某家公司是否有偷税漏税行为。否则，税务审计员必须从头到尾查看成千上万的 PDF 文档。在公共数据库生命周期管理中，对一大类分布式监测数据，如环境监测、道路视频监控、消防、特种设备运行监测等，数据当局在 Hadoop 和非 Hadoop 环境下对其数据归档时必须考虑可达到的压缩程度。就目前而言，LZO、Gzip 和 RainStor 是较流行的高效数据压缩技术。

4) 管理实时数据流的生命周期。对公共数据库中的高速实时数据流，数据当局必须明确其保存价值，即是否需要永久保存这些数据流。这种高速数据流往往由机器 (含传感器) 自动产生，当机器数据产生异常行为时需要存储该异常事件发生前后的每一个读数。例如，网络监测系统异常事件的数据流获取中公安局网监大队需要确定保存在内存中的数据应该是多少，可供选择的方案为内存中保存 2 小时的 NetFlow 记录并每隔一分钟将记录保存于硬盘一次，以备历史分析之用。

5) 保留适量社会团体和商业机构数据交互记录，并支持电子证据展示。随着公民社会的日益临近，公共部门将逐步退出某些领域，如科学技术奖励评审、职称评审、信用管理、社会公证等，这些领域的公共数据管理也将伴随公共事务移交给社团组织，如工会、政党、行业协会、公众企业 (如大学、医院、BAT、华为技术公司、国家电网公司等)，相应的公共数据资源也将转移到这些社团组织和商业机构。但是，法律仍然赋予了这些组织一定的公共管理职能，如邮政公司有义务承担偏远农村地区物流、信函投递等。伴随这些非政府组织的公共服务职能的公共数据管理也必须支持电子证据展示。例如，人民银行数据当局要求

商业金融机构保留通过社交网站与顾客的交流记录, 这些数据有的还涉及数据业务承包商的控制。因此, 公共数据生命周期管理必须考虑到司法、仲裁等公共事务对这些外围数据的电子证据需求, 并制定相应的政策或法律。

6) 按照法律和制度要求, 定期处置不再需要的公共数据。很多公共部门认为永久保留公共数据是对法律和公众质询最好的应对之策。但是, 任何一项数据资源的持有都将产生财务成本, 而且法律对违法案件都有一定的追溯期限, 实际上已经将一部分公共数据确定为负资产。按照法规, 对这部分丧失电子证据作用的公共数据资源做出适当处置是公共数据生命周期管理的重要环节。

4 电子证据的关联

电子证据是存在于虚拟空间的证据, 它离不开由电子设备和信息技术所营造的特殊环境, 这种环境的特殊性决定了它与传统人证、物证相比显然不同。首先, 电子证据的虚拟空间通常不是真实的物, 而是由某种信号的方式存储的信息[17]。其次, 电子证据所存在的虚拟空间非常繁杂, 无论是网络、云盘, 还是光盘、硬盘、U 盘等电子信息空间, 人们都无法触及。通常这些虚拟空间和案件事实存在的物理空间须经过某种转换才能建立相应的联系。这种联系即为内容关联性与载体关联性。内容关联性即电子证据的数据内容与案件事实之间的关联性, 载体关联性即电子证据的信息载体同当事人或其他诉讼参与人之间的关联性。因此, 法庭是否必须对电子证据作出双重的关联性判断取决于这种双联性原理。只要缺少对任一关联性的认定或是对任一种认定的结论是否定的, 都会导致法庭不采纳电子证据。

4.1 电子证据的内容关联性

在公共数据库的电子证据的关联性研究方面可供借鉴的成果还很少, 问题也更加复杂, 而且具有很强的实务性。从电子证据的内容关联方面讲, 就涉及到高度复杂的数据智能串并分析。例如, 在公共安全数据库中, 为了提高破案的成功率, 必须要对案件库进行智能案件串并分析, 即以人、事件、物、机构、地点等要素对各业务部门综合应用加以涵盖和抽象, 在此基础上提出数据的关联串并要求, 并最终形成数据链, 作为证据提供给检察院、法庭和仲裁庭等。由于这些电子数据将作为证据的一部分, 其本身也涉及法律适定性。因此, 作为电子证据的公共数据链必须遵从严格的逻辑关系, 包括要素逻辑和时序逻辑[17]。我们以公安数据库为例, 阐述公共数据库的电子证据的串并逻辑。该逻辑体系主要有以下几个方面:

1) 人员要素串并逻辑: 人员要素信息是公共安全工作的基础, 它涉及公安业务的方方面面。各业务部门数据几乎都涉及到人员信息, 包括常住人口、暂住人口、流动人口、关押人员、犯罪嫌疑人等, 总计 40 多种, 涉及的部门众多, 包括治安、交管、刑侦、禁毒、监管、外管等。

2) 物品要素串并逻辑: 在公安业务处理过程中, 凡涉及物品的业务数据, 都是物品要素关联的范围, 包括证件、枪支、爆炸物品、机动车、涉案物品等。

3) 事件要素串并逻辑: 事件要素关联的范畴包括公安业务中凡是跟案件有关的处理过程和数据, 如治安案件、案事件笔录等。

4) 机构要素串并逻辑: 组织要素包含范围涵盖公安日常业务管理涉及的机构和涉案机构, 包括旅馆、特业机构、涉枪机构、涉爆机构、房屋出租机构、单位犯罪等。

5) 地点要素串并逻辑: 地点要素是快速反应、快速定位的关键, 凡是涉及地点或地址的数据都是地点要素关联的范畴。

6) 时序逻辑: 公共安全案件必须反映整个作案过程, 时间是整过演化过程必不可少的参量, 因此电子证据数据之间必须具备严格的时序逻辑关联。

4.2 电子证据的载体关联性

对于电子证据的载体关联性, 司法实践中几乎找不到完全相同的两起案件, 从结构上来说, 任何案件都是由人、事、物、时、空构成的, 对应于虚拟空间的身份、行为、介质、时间、地址。也就是说, 在司法实务中, 法庭要通过确认涉案信息载体的身份、行为、介质、时间、地址关联性, 将物理空间与虚拟空间的案件事实关联起来。

1) 身份关联性。在虚拟空间中, 人的关联性体现在人的身份上, 主要表现为各类电子账号。在具体案件事实中, 需要能够证明案件所涉及的电子账号归当事人或其他诉讼参与人所属或所用。这种身份关联性的构建必须排除涉案电子账号共有、共用或者案外人使用、冒用的情况。这实际上是证明当事人或其他诉讼参与人就是虚拟空间中以某个特定身份行事之人。2) 行为关联性。案件事实涉及的各种法律责任一般发生在物理空间, 但在虚拟空间中的案件事实则需确认当事人或其他诉讼参与人是否实施相关行为, 比如

是否收发了一条短信、一封邮件，是否修改了某个文档或下载了某个网页等。这些行为将影响对当事人等主体法律责任的最终认定。3) 介质关联性。电子证据需由硬盘之类的电子介质承载，因此需要确认此类介质同当事人或其他诉讼参与人的关系，若存在该电子介质为当事人或其他诉讼参与人共有或共用的情况，那么需确立电子介质中的数据同当事人或其他诉讼参与人之间的对应关系。4) 时间关联性。物理空间的时间与机器时间具有一定的对应关系，但又不完全一致[18]，虚拟空间的时间通常是机器时间。时间关联性就是要确立物理时间与机器时间是否一致以及其对应关系如何，从而确定在案发时间人的行为产生的相应电子证据。在司法实务中，时间是定案的关键要素，如果出现物理时间与机器时间不同，就会带来时间关联性问题。5) 地址关联性。虚拟空间有独特的地址概念。大多数电子证据产生后都带有地址信息。这就要求确认这些地址信息与当事人或其他诉讼参与人之间的关系。

上述身份、行为、介质、时间、地址的关联性均立足于虚拟空间，共同构成了电子证据的载体关联性。在具体案件中，只有存在争议的关联性问题才具有实际意义，这正是电子证据关联性独特之处。

5 电子证据区块链系统原型系统概述——公安执法记录区块链为例

从公安民警接警、处警、立案、破案、结案到检察院起诉和法院判决过程中，电子证据的证据作用越来越大。其中既包括处警过程中自动数据采集系统获取的电子数据 (如道路视频监控、楼宇安防视频监控等)、警用执法记录仪及平台获取的处警过程电子数据、办案大厅出入门禁 (A/B 门)、电子手环数据、问询笔录电子扫描件和音/视频电子媒介及其承载的电子数据，也包括立案后侦查过程中获取的人证、物证的相关电子数据。为了提高整个办案过程中形成的电子证据的可信度，本节提出了一个去中心化的警用电子证据区块链系统原型。该电子证据系统的目的是形成不可更改并且保持所有办案环节所呈现的电子证据内容关联一致性，称之为证据共识。区块链的形成如下：

时间戳 1

内容：报警记录，处警任务单，警用车辆号牌，执法记录仪编号，执法记录仪获取的音视频电子资料，归档的电子材料。

签名：警官 A 的签名、警官 B 的签名、警官 C 的签名……(至少 3 名警官签名)；现场目击证人签名，现场被问询人签名。

说明：假定警官 A 负责电子证据采集，他将发布警情记录表，同时发给每一个参与人 (签名者)。由于警情是会变化的，所以后续每个时间戳下的电子证据记录的事项是会改变的，因此不一定每次都是由警官 A 负责电子证据发布，甚至有新的警官或涉案人员、证人加入电子证据区块链，故可有新的警官进行电子证据发布。执法记录仪必须符合公安部的认证、在有效使用期限内并且各项功能正常，执法记录仪的电子证据提取有自动提取和和手工提取 2 种方式。电子证据的自动提取或手工提取必须使用执法记录仪自带的操作系统，多余的电子材料也只能系统定期自动覆盖，不能进行人工操作 (合法的执法记录仪必须有此项功能)。智能型执法记录仪截取的电子证据需满足必要的清晰度并具有证据意义，手工提取电子证据必须考虑电子证据的关联性。所获取的电子证据依法存档并进入办案大厅信息平台中的对应区块。更重要的一点是，为了保证电子证据的可信度，要让每个参与者 (签名者) 独立确认电子证据区块链的变化，前提是 2/3 以上参与者同意区块链上删改或新增的内容。为什么要这样呢？为了达到证据共识，需要大多数人承认电子数据的真实性。但是，为什么区块链系统不要求所有的人都要签名呢？因为这会让少数缺席或拒签者危及整个共识的形成，尤其是可能的涉案嫌疑人的抵赖行为影响电子证据的形成，也可以避免个别参与者因故不能及时签名或警员回避而导致案件续侦过程中电子证据的伴随生成。同时，这样做会避免在电子证据中给予某个人太大的权利。之所以需要 2/3 多数而不是过半多数签名才有效，是因为如果只是过半多数签名有效的话，按 Byzantine 将军算法，可能会形成两个相互矛盾的电子证据版本。

……

时间戳 n

内容：办案大厅门禁系统电子记录 (A/B 门)、办案区涉案人员电子手环记录、涉案人员身份查证 (身份证、免冠正面照、生物特征等)，问询影响录音电子数据和笔录签名扫描件等。

签名：警官 E 的签名、警官 F 的签名、警官 G 的签名……警官 A 的签名、警官 B 的签名、警官 C 的签名、被询问人签名、现场目击证人签名，现场被问询人签名。

说明: 设置办案大厅是公安执法规范化的基本要求。除非特殊情况, 公安民警处警过程中, 必须将涉案人员带回规范设置的办案大厅进行问询和进一步取证。办案大厅的功能多样化, 一方面可以有效避免警务人员滥用职权、刑讯逼供、伪造证据等, 以此保护涉案人员的公民权利, 也可以有效保护办案警务人员的人身安全、被诬陷; 另一方面, 也是更为重要的方面, 通过办案大厅可以有利于证据的获取和提高证据的采信度, 其中也包括电子证据的获取、固定和保存, 以此提高电子证据的可信性。因此, 办案大厅设置及软硬件设计必须有利于提高电子证据的可信度。目前已有的办案大厅基本采用了传统的中心化数据库和加密措施, 但是这种的中心化数据库和加密措施并不能保证电子证据材料的不可更改性, 仍然有可能被篡改或删除。为此, 本文在办案大厅的各个环节的电子证据取证过程中运用区块链技术。

……

时间戳 $n+m$

内容: 立案卷宗 (不予立案归档), 立案程序电子记录, 涉案人员拘押程序电子记录等。

签名: 检察官 A、检察官 B、警官 H、警官 I……警官 E 的签名、警官 F 的签名、警官 G 的签名、警官 A 的签名、警官 B 的签名、警官 C 的签名、被询问人签名、现场目击证人签名、现场被问询人签名……。

说明: 经过在办案大厅的初步调查, 视情况进行下一步程序。若警情达不到立案条件, 则办理归档手续, 释放涉案人员; 若已经具备立案条件, 则启动立案程序。立案程序中包括公安部门内部程序和外部程序。外部程序中包括向检察院提报涉案人员拘押、批捕申请和向法院提报的财产保全、相关场所查封申请等。

时间戳 $n+m+1$

内容: 侦查阶段, 案件侦查中人、事、物关联和取证, 获取更多的证据, 并产生相应的电子证据材料。电子证据的关联包括高危人群特征数据智能比对、刑满释放及在逃人员数据比对、挂失物品登记数据智能比对、车辆及驾乘人员登记数据比对、车辆维修记录比对、宾馆及娱乐场所消费记录数据比对、多案智能串并分析等结果的电子记录。其中部分电子证据关联分析须经有经验的警务人员或聘用的专业机构及个人进行人工干预, 则须加入技术检定人员资格认定和参与过程记录。

签名: 侦查警官甲、侦查警官乙、侦查警官丙……同案嫌疑人 A、同案嫌疑人 B、同案嫌疑人 C……证人 A、证人 B、证人 C……委托证物检定人 A、委托证物检定人 B、委托证物检定人 C……证言证物登记人 A、证言证物登记人 B、证言证物登记人 C……检察官 A、检察官 B……警官 H、警官 I……警官 E 的签名、警官 F 的签名、警官 G 的签名、警官 A 的签名、警官 B 的签名、警官 C 的签名、被询问人签名、现场目击证人签名, 现场被问询人签名……

时间戳 $n+m+2$

内容: 结案申请和批准。提交结案申请表、结案材料、卷宗副本交接等。若结案申请被批准, 则建立卷宗、办理结案手续; 若结案申请被退回, 则进入补充侦查阶段。

签名: 检察官甲、检察官乙……侦查警官甲、侦查警官乙、侦查警官丙……同案嫌疑人 A、同案嫌疑人 B、同案嫌疑人 C……证人 A、证人 B、证人 C……委托证物检定人 A、委托证物检定人 B、委托证物检定人 C……证言证物登记人 A、证言证物登记人 B、证言证物登记人 C……检察官 A、检察官 B……警官 H、警官 I……警官 E 的签名、警官的签名、警官 G 的签名、警官 A 的签名、警官 B 的签名、警官 C 的签名、被询问人签名、现场目击证人签名, 现场被问询人签名……

时间戳 $n+m+3$

内容: 补充侦查阶段, 案件侦查中人、事、物补充关联和补充取证, 高危人群特征智能比对校准和更正, 多案智能串并分析等, 获取更多的证据, 完善证据链, 并产生相应的电子证据材料。

签名: 侦查警官金、侦查警官木、侦查警官水……新证人甲、新证人乙……侦查警官甲、侦查警官乙、侦查警官丙……同案嫌疑人 A、同案嫌疑人 B、同案嫌疑人 C……证人 A、证人 B、证人 C……委托证物检定人 A、委托证物检定人 B、委托证物检定人 C……证言证物登记人 A、证言证物登记人 B、证言证物登记人 C……检察官 A、检察官 B……警官 H、警官 I……警官 E 的签名、警官 F 的签名、警官 G 的签名、警官 A 的签名、警官 B 的签名、警官 C 的签名、被询问人签名、现场目击证人签名, 现场被问询人签名……

时间戳 $n+m+4$

……

6　结束语

本文建立了中国公共数据库电子证据系统的

区块链模型，较好地解决了公共部门运用公共数据库自动提取数据并生成不可更改或删除的电子证据的关键技术，并通过电子证据取证系统有效防范取证过程及证据转移过程中可能发生的电子证据改变，提出运用数据生命周期管理持续改进公共数据库电子证据的证据效力。更进一步，从法律的角度，讨论了电子证据运用环节的电子证据关联性，初步给出了载体关联、内容关联、案件串并的一般原则。

由于公共数据治理的复杂性，在某些方面研究尚不充分。在基于中国公共数据库的电子证据关联性分析中，必须把司法实务中的专业领域知识与电子证据信息相结合，构成“人-机”结合的时空关联、载体关联和内容关联的知识自动化系统模型。

参考文献：

[1] LEE Y W, PIPINO L L, FUNK J D, et al. Journey to data quality[M]. Cambridge, Massachusetts: The MIT Press, 2006.

[2] 袁勇, 王飞跃. 区块链技术发展现状与展望 [J]. 自动化学报, 2016, 42(4): 481–494.
YUAN Yong, WANG Feiyue. Blockchain: the state of the art and future trends[J]. Acta automatica sinica, 2016, 42(4): 481–494.

[3] 张宁, 王毅, 康重庆, 等. 能源互联网中的区块链技术: 研究框架与典型应用初探 [J]. 中国电机工程学报, 2016, 36(15): 4011– 4022.
ZHANG Ning, WANG Yi, KANG Chongqing, et al. Blockchain technique in the energy internet: preliminary research framework and typical application[J]. Proceedings of CSEE, 2016, 36(15): 4011– 4022.

[4] EDWARD MEINERT, ABRAR ALTURKISTANI, KIMBERLEY A FOLEY, et al. Blockchain implementation in health care: protocol for a systematic review[J]. JMIR research protocols, 2019, 8(2): 153–159.

[5] GUO Rui, SHI Huixian, ZHAO Qinglan, et al. Secure attribute-based signature scheme with multiple authorities for blockchain in electronic health records systems[J]. IEEE access, 2018, 6: 11676–11686.

[6] 王桂强. 电子物证检验 [J]. 刑事技术, 2003(4): 3–7.
WANG Guiqiang. Examination of electronic evidence[J]. Forensic science and technology, 2003(4): 3–7.

[7] 刘品新. 电子证据的关联性 [J]. 法学研究, 2016, 38(6): 175–190.
LIU Pinxin. Relevance of electronic evidence[J]. Chinese journal of law, 2016, 38(6): 175–190.

[8] MA Zhaofeng, HUANG Weihua, GAO Hongmin. Secure DRM scheme based on blockchain with high credibility[J]. Chinese journal of electronics, 2018, 27(5): 1025–1036.

[9] HAWLITSCHEK FLORIA, NOTHEISEN BENEDIKT, TEUBNER TIMM. The limits of trust-free systems: A literature review on blockchain technology and trust in the sharing economy[J]. Electronic commerce research and applications, 2018, 29: 50–63.

[10] YUAN Hang, ZHANG Shibin. Study on design and application of electronic evidence preservation program[C]//Proceedings of 2011 International Conference on Internet Technology and Applications. Wuhan: IEEE, 2011: 1–4.

[11] 蔡维德, 郁莲, 王荣, 等. 基于区块链的应用系统开发方法研究 [J]. 软件学报, 2017, 28(6): 1474–1487.
TSAI W T, YU Lian, WANG Rong, et al. Blockchain application development techniques[J]. Journal of software, 2017, 28(6): 1474–1487.

[12] TIAN Zhihong, LI Mohan, QIU Meikang, et al. Block-DEF: A secure digital evidence framework using blockchain[J]. Information sciences, 2019, 491: 151–165.

[13] 刘文奇. 复杂网络上的公共数据演化博弈与数据质量控制 [J]. 中国科学(信息科学), 2016, 46(11): 1569–1590.
LIU Wenqi. Public data evolution games on complex networks and data quality control[J]. Scientia sinica informationis, 2016, 46(11): 1569–1590.

[14] 刘文奇. 中国公共数据库数据质量控制模型体系及实证 [J]. 中国科学(信息科学), 2014, 44(7): 836–856.
LIU Wenqi. Modeling data quality control system for Chinese public database and its empirical analysis[J]. Scientia sinica informationis, 2014, 44(7): 836–856.

[15] ELISA NOE, YANG Longzhi, CHAO Fei, et al. A framework of blockchain-based secure and privacy-preserving E-government system[J]. Wireless networks, 2018: 1–11.

[16] 王继业, 高灵超, 董爱强, 等. 基于区块链的数据安全共享网络体系研究 [J]. 计算机研究与发展, 2017, 54(4): 742–749.
WANG Jiye, GAO Lingchao, DONG Aiqiang, et al. Block chain based data security sharing network architecture research[J]. Journal of computer research and development, 2017, 54(4): 742–749.

[17] 王电. 公安信息化概论 [M]. 北京: 清华大学出版社, 2011.
WANG Dian. An introduction to public security information[M]. Beijing: Tsinghua University Press, 2011.

[18] 费敏锐, 熊南, 李韬. 网络化系统时钟同步算法 [J]. 中国科学(信息科学), 2016, 46(11): 1527–1541.
FEI Minrui, XIONG Nan, LI Tao. Clock synchronization algorithms for networked systems[J]. Scientia sinica informationis, 2016, 46(11): 1527–1541.

作者简介:

李萌, 女, 1994 年生, 硕士研究生, 主要研究方向为公共数据治理、数据质量控制。

刘文奇, 男, 1965 年生, 教授, 博士生导师, 主要研究方向为数据博弈、数据质量控制、复杂系统建模。主持国家级项目多项。发表学术论文 20 余篇。

米允龙, 男, 1986 年生, 博士研究生, 主要研究方向为机器学习、数据挖掘。发表学术论文 6 篇。

IEEE Conference on Computer Vision and Pattern Recognition （CVPR 2020）

The CVPR 2020 is dedicated to computer vision and pattern recognition including color detection, tracking, motion, object recognition, stereo and object detection.

The CVPR 2020 covers topics such as: Segmentation and Grouping, Motion and Tracking, Human Recognition, Shape-from-X, Stereo and Structure from Motion, Color and Texture, Illumination and Reflectance Modeling, Image-Based Modeling, Sensors, Shape Representation and Matching, Computational Photography and Video, Early and Biologically-Inspired Vision, Video Analysis and Event Recognition, Optimization Methods, Face and Gesture Analysis, Video Surveillance, Scene Understanding, Image and Video Retrieval, Medical Image Analysis, Vision for Robotics, Vision for Graphics, Statistical Methods and Learning, Applications of Computer Vision, Document Analysis, Object Recognition/Detection/Categorization.

会议日期: 14-19 June 2020

会议地点: Washington State Convention Center, Seattle, WA, United States

网站: https://www.clocate.com/conference/cvpr/14111/