

# Research on the Textual Measurement of Electronic File in Public Security Cases—Taking Telecommunication Fraud Cases as an Example

Jiawei Lu, Chong Tian, Bowen Li, Jiuyi Shen, Jiawei Hu, Lin Guan\*

Jiangsu Police Institute, Nanjing Jiangsu  
Email: \*17590324@qq.com

Received: Jun. 15<sup>th</sup>, 2020; accepted: Jun. 22<sup>nd</sup>, 2020; published: Jun. 29<sup>th</sup>, 2020

---

## Abstract

With the development of big data technology in recent years, electronic files have become popular in public security organs. Big data mining for electronic files has yet to be promoted. This article uses text measurement methods, based on the dimensions of the time and place of the telecommunications fraud case, the characteristics of the case, etc., through word frequency analysis, geographic information visualization tools and other methods to carry out research on electronic files related to the trial documents of telecommunications fraud cases, with the aim of discovering the common characteristics and development rules of the case. The results of the text analysis are verified with reference to the case statistics of the public security organs, so as to provide a reference for the public security organs to combat telecommunications fraud.

## Keywords

Electronic Dossiers, Trial Documents, Telecommunications Fraud, Text Measurement

---

# 公安案件电子卷宗文本计量研究——以电信诈骗案件为例

陆家炜, 田 翀, 李博文, 沈久一, 胡佳伟, 关 琳\*

江苏警官学院, 江苏 南京  
Email: \*17590324@qq.com

---

\*通讯作者。

文章引用: 陆家炜, 田翀, 李博文, 沈久一, 胡佳伟, 关琳. 公安案件电子卷宗文本计量研究——以电信诈骗案件为例[J]. 数据挖掘, 2020, 10(3): 221-228. DOI: 10.12677/hjdm.2020.103023

收稿日期：2020年6月15日；录用日期：2020年6月22日；发布日期：2020年6月29日

## 摘要

近年来随着大数据技术的发展，电子卷宗已在公安机关普及。针对电子卷宗的大数据挖掘尚待推进。本文利用文本计量方法，基于电信诈骗案件的发案时间地点、案件特点等维度，通过词频分析、地理信息可视化工具等方法，对涉及电信诈骗案件审判文书的电子卷宗开展研究，旨在发现此类案件的共性特征以及发展规律，并参照公安机关案件统计数据对文本分析的结果加以验证，以此为公安机关打击电信诈骗犯罪提供参考。

## 关键词

电子卷宗，审判文书，电信诈骗，文本计量

Copyright © 2020 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

电子卷宗的推广和普及是公安事业改革的新起点。大数据时代电子卷宗对案件的保存、记录和传输具有重要意义。娄永涛指出电子卷宗的广泛应用为司法办案信息化和大数据化创造了先决条件[1]。电子卷宗是指在案件受理前或者案件受理过程中，将原始纸质案卷材料依托数字影像技术、文字识别技术、数据库技术等媒介技术制作而成的具有特定格式的电子文档和相关电子数据。公安机关通过卷宗的电子化，将纸质化的涉及到公安案件的卷宗上传到案件信息公开系统等公安机关专用系统之上，王雪认为这一举措对公安机关记录和保存案件起到重要的作用，并有助于公安机关联系、侦破串并案[2]。

随着大数据时代的到来，数据的流动和共享已经成了一把“双刃剑”，利用大数据实施电信诈骗的行为日益猖獗。诈骗分子利用数据获取的便利，衍生出了许多实施诈骗的套路，如贷款诈骗、冒充熟人诈骗、利用培养感情诈骗、推销商品(保健品)诈骗等多种符合被害人需求的诈骗套路。网络的全球化更是滋生了跨国跨境的电信网络诈骗，然而由于国家之间法律、文化的差异和地域对跨境办案的限制，使得打击跨境电信网络诈骗难上加难，这类犯罪也日益猖獗，难以得到有效遏制。

电信诈骗案件的侦办工作往往以受害人报警为触发点，公安机关传统的侦办方式更加侧重于对单一或单系列案件的研判，缺乏对一定时期内此类案件整体发案情况的掌握。这种犯罪研究方式难以从宏观层面发现电信诈骗案件的整体规律[3]。加之电信网络诈骗案件作案手段种类多、变化快，也加大了提前预防的难度，使得公安机关处于疲于应付的局面，亟待探索一种新的案件研判预防思路[4]。本文拟采用文本计量方法，对此类案件的电子卷宗开展研究，旨在运用定量分析工具发现电信诈骗案件的规律，为公安机关常态化预防和打击电信网络诈骗案件提供新的思路。

文本计量分析法于1911年由俄国化学家瓦尔金所开创，1969年由英国学者普理查确定了文本(献)计量学这一术语[5]。文本计量分析法是以文本的各方面属性为基础，通过数学和统计学方法，把文本的特征和体系作为研究对象，研究文本的变化规律、分布结构和数量关系。文本分析法是将非结构化的文本信息转化为结构化的定量信息，进而实现了量化分析，此方法具有客观、定量和易于比较的优势。近

年来,文本分析法在规划涉及领域应用前景广泛,但到目前为止,将其应用于公安犯罪领域研究还比较鲜见。

## 2. 研究样本与研究方法

### (一) 研究样本

本研究的数据来自于中国裁判文书网(<http://wenshu.court.gov.cn>),我们将检索方式设置为高级检索,检索第一关键词确定为“电信诈骗”,第二关键词选择为“网络诈骗”,进行精确检索,截止至2020年1月1日,通过上述渠道一共搜集到了2499篇电信诈骗案件的裁判文书。通过分析对比,从裁判文书的数量分布上来看,自2017年开始,电信网络诈骗的发案量明显有大幅度增长,相比2016年增长了10倍之多,自此电信网络诈骗案件的发案率居高不下近年来持续增长,显然公安机关还未找出有效措施来遏制这一趋势。具体发案量变动情况见图1。

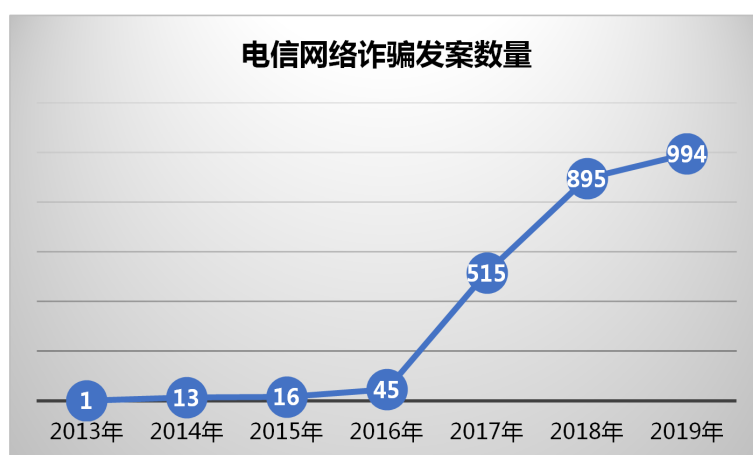


Figure 1. The number of fraud cases in telecommunications networks  
图1. 电信网络诈骗发案数量图

### (二) 研究方法

本文采用文本计量方法对电信诈骗案件的裁判文书的正文内容进行词频统计和计量分析。词频分析是一种内容分析方法,根据词频统计的规则,区分出高频词汇和低频词汇,并采用“可视化”的方法,通过词云、数据地图等文本分析的工具,筛选出具有代表性的高频词汇,从多个维度加以分析。如果将词频以统计表的方式列出,可以直观感受到高频词之间的频次差异[6]。一个单词或名词或短语出现频率越高,就越能体现出整篇文书的主题,高频词反映出了电信诈骗案件中的特点、诈骗话术常用词等。

“词云”这个概念由美国西北大学新闻学副教授、新媒体专业主任里奇·戈登(Rich Gordon)提出。“词云”是数据可视化的一种形式,就是对文本中出现频率较高的关键词生成一副图像,予以视觉上的突出,形成“关键词云层”或“关键词渲染”。从而使得浏览者只要一眼就能领略文本的重点信息和主旨。

数据地图是一种图形化的数据表现的工具。当数据中含有地理位置信息的时候,可以使用数据地图工具来分析和展示与地名相关的数据,把这些数据投射到实际的地图上,使浏览者可以更加清晰直观的看出数据的实际地理分布情况。本文采用的“地图慧”,就是一款专业的数据地图制作工具。

本研究还使用了武汉大学沈阳教授团队研发的文本采集分析工具(Rost),可以自动分析和采集裁判文书网上的有效内容,对电信诈骗案件的裁判文书中的重点信息加以分析,再结合网络上对于电信诈骗案

件特点的研究分析，以此评估用文本分析工具研究公安电子卷宗的准确性和可行性。

### 3. 利用个人信息泄露是诈骗实施的主要手段

从搜集到的这 2499 篇电信诈骗案件的裁判文书中提取高频词汇，可以发现电信诈骗案件的常规作案手段，并初步总结电信网络诈骗案件的基本特点。在这些高频词汇中，“公司”(24,702 次)、“账户”(18,497 次)、“手机”(17,987 次)、“银行”(12,042 次)等词语出现频率较高，并且具有一定的代表性。比如“手机”、“电话”词等表明了电信诈骗案件主要是通过手机、固定电话、网络等通信工具和现代网银技术实施的非接触式的诈骗犯罪，这些犯罪类型繁多，包含虚构“紧急情况”诈骗、互联网购物诈骗等各种各样的手段，通常来说在犯罪人特意针对下，降低受害人安全意识，使他们难以防备，这样往往容易中计，给犯罪人有隙可乘。具体经筛选的高频词汇见下图 2。



Figure 2. High-frequency words in telecom fraud referee documents  
图 2. 电信诈骗裁判文书高频词

在高频词中“手机”一词的出现次数多达 17,987 次，手机关联到了移动支付。如今，“支付宝”、“微信支付”、“网上银行”的便利使得支付转账无需再像从前那样到实体银行网点去实施转账汇款等操作，只需要在手机上绑定银行卡账号、输入密码等一系列操作，在很短的时间内就可以完成转账汇款。支付方式如此便利使得受害者甚至没有犹豫的时间，财产损失瞬间发生。在侦办电信诈骗案件过程中总结出的电信诈骗关键词，也有类似“只需手机就能完成”、“移动支付转账”等关键词，这些都是出现在犯罪分子套路话术当中的关键词。他们通过告诉受害者“xx 赚钱”的便利，只需要手机操作就能获利，让受害者感受到“xx 赚钱”的成本低、操作简单，就会让人轻易相信并愿意尝试犯罪分子所谓的“赚钱理财”，于是结果可想而知。本研究得出的结果与此类案件侦办经验相符。

“信息”一词的出现次数也多达 16,320 次，“个人信息”出现次数多达 4331 次，从这可以看出个人信息的泄露很大程度上为诈骗分子实施电信诈骗提供了便利。例如 2019 年最热门的新词“杀猪盘”，这便是一种从泰国、缅甸等东南亚国家流传过来的新型诈骗方式。诈骗分子利用网络社交软件与受害者“结识”并“培养感情”，业内的行话便把这种“培养感情”称为“养猪”，以谈恋爱的手段获取受害

者的信任，用甜言蜜语哄骗受害者便是“喂猪饲料”，当感情培养到一定阶段的时候，诈骗分子便会向受害者介绍网络博彩、网络贷款或是投资理财的手机软件，引诱受害者进入骗局，不断的投入资金导致一步一步越陷越深，直至倾家荡产。

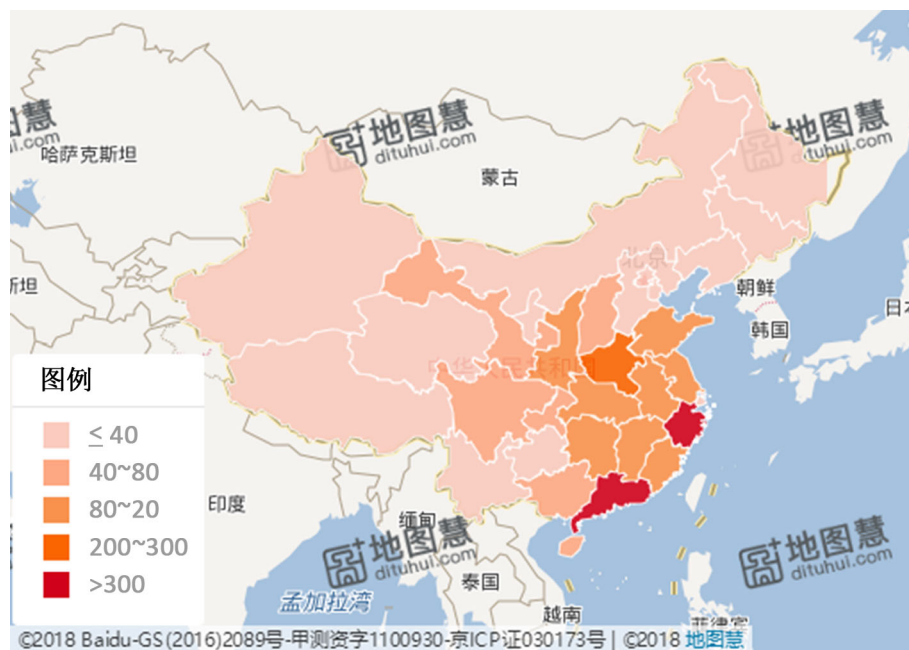
诈骗分子利用身份信息寻找零星作案对象。2019年9月19日召开的“网络婚恋平台诈骗防范治理研讨会”会上发布的数据显示：2019年1月到7月，百合佳缘集团旗下的世纪佳缘APP收到的涉及侵财类诈骗案件举报数多达1482例。公安部门的数据显示，2019年1月到8月，网络交友诱导赌博、投资诈骗共造成群众财产损失达38.8亿元人民币，占到全部电信网络诈骗损失的21.3%，在所有的诈骗类型中造成的损失数最大。诈骗分子利用这些网络社交交友软件上用户注册时填写的个人信息，有些隐私信息遭到窃取，有些个人信息甚至可以公开获取。通过获取这些个人信息，诈骗分子就可以筛选出诈骗对象，也就是这批有感情需求的，年龄往往在25岁到40岁之间的特定人群，而这批人群往往又是一些有稳定收入的或是有一定经济实力的单身男女，禁不住感情的诱惑，就心甘情愿把钱投入到网络博彩、网络理财软件。诈骗分子还会利用受害者的贪婪心理，当受害者参与网络博彩赔钱的时候，鼓励受害者使用高额利息的网络借贷软件，有些企图“一记翻本”的受害者就会陷入一个更深的圈套。有些人知道参与网络博彩的违法性质，在遭遇诈骗之后不选择报警，甚至会听信诈骗分子以交纳“保证金”来“解冻”先前投入资金的借口，再被诈骗分子捞走一笔财产。这样的结局往往就是倾家荡产，甚至可能是家破人亡。这类电信诈骗案件涉及金额巨大，一定程度上也导致了受害者自杀。

诈骗分子利用他人信息包装身份实施诈骗。“照片”一词的出现次数高达4251次，“号码”一次的出现次数也高达3548次。诈骗分子实施诈骗利用的手机号码、照片等，正是我们平时生活中被泄露的个人信息。诈骗分子通过搜集我们发布在社交平台上的日常生活照片等，组合成系列照，打造了一个全新的虚假身份，在进行感情诈骗的时候，给诈骗对象发送这些照片，照片中都是同一个人在不同生活状态的记录，看似就是正常的日常生活照片，诈骗对象心里就自动形成了一个虚假的帅哥美女或者是成功人士的形象，降低了诈骗对象的心理防线，更容易骗取诈骗对象的信任。而获取这些个人信息的渠道，可以是诈骗分子在社交平台上窃取的，甚至这种个人信息已经被作为商品打包出售。曾有研究指出，在淘宝网这样的大型电商购物平台上隐藏着一些出售类似个人信息的店铺，只需支付一笔小小的费用，就可以得到一个人的个人信息包括一系列的证件照片和日常生活照片。这甚至已经成为了一种地下产业，隐藏在我们日常使用的电商购物平台里。只不过像我们平常购物时搜索关键词那样是无法搜索到这些店铺的，要输入一些字符代码暗号或者行业黑话、暗语作为搜索关键字，这些出售个人信息的不法店铺才会出现。

诈骗分子利用已泄露的身份信息精准分人群实施系列诈骗。例如，由于学生个人信息的泄露，诈骗分子就会冒充高校招生办老师、学校辅导员等进行诈骗，由于很多学生还未正式踏入社会，警惕性相对较低，并且对能够准确报出自己名字、学号等信息的“学校老师”深信不疑，诈骗分子往往就很容易得逞。这类型案件中受害者的自杀率也相当高，很多学生家庭条件一般，家里人辛苦攒下供孩子上学的钱就这么被骗子轻而易举地骗走，一部分学生十分懊恼后悔又不敢和家里人说，迫于压力最终走上了自杀的道路。

#### 4. 新型诈骗手段大多由境外传入

近年来电信诈骗案件主要在浙江省(243)、广东省(319)、河南省(243)、江苏省(175)、福建省(174)等地区较为集中，发案数量较多，这些地方的公安机关所面临的问题较为严峻；而贵州省(8)、青海省(5)、西藏自治区(4)、宁夏回族自治区(2)的电信诈骗案件较少，表现出良好的形势。针对这一现象，有关地区的公安机关应当共享相关的信息技术和情报，来共同应对这些电信网络诈骗案件。地区发案数量见下图3。



**Figure 3.** The location of the telecommunications fraud  
**图 3.** 电信诈骗案发地

根据利用“地图慧”工具绘制出的数据地图，可以清楚直观的看出，我国东南沿海地区被标记为了红色，往内陆地区延伸，逐步由红色降至橙色再降至淡红色。由此可以得出，我国东南沿海经济发达的地区电信诈骗案件的发案量占比较大。其中广东、福建、浙江、江苏四地的电信诈骗发案量达到全国电信诈骗案件总量的 37%。电信诈骗案件的发案量从我国东南沿海地区向西北地区递减，呈现出从海外向我国沿海到内陆扩散的态势。这一结论较好的印证了，相关学者的前期研究结论，即很多的诈骗方式是从海外流传到中国大陆，从沿海地区逐步向内陆地区蔓延的[7] [8]。除了上述提到的“杀猪盘”，是从泰国、缅甸等东南亚国家流传过来，先进入台湾地区，再从台湾地区传入大陆，其服务器和基地都是设立在东南亚国家的，另一种“涉案类”电信诈骗的方式也同样由台湾地区传入大陆[9]。

所谓“涉案类”电信诈骗，是指诈骗分子冒充公、检、法等国家机关的工作人员，利用伪基站群发短信和来电任意显示的功能，声称受害人涉嫌洗钱、包裹涉毒、银行卡有巨额消费异常、电话欠费、医保卡信息泄露或者大量购买药物涉嫌套现犯罪等作为借口，诱骗受害人把资金转入诈骗分子提供的所谓“银行安全账户”，当这批资金经过审查后方可转出。很多人为了自证清白，不怕接受查验，主动配合所谓“公检法机关”查案，于是就落入了诈骗分子的圈套中去。

这些类型的电信诈骗犯罪团伙，其基地往往设立在境外，或是已经从境外传入台湾地区，再通过台湾地区设立总部，利用大陆人员警惕意识尚且薄弱，对我国东南沿海地区的人员实施诈骗，再一步步地往内陆延伸。由于其服务器和总部不在境内，不仅直接导致了受害者的被骗财产难以被追回，更是大大加深了我国警方打击跨境电信诈骗犯罪的难度。

### 5. 文本分析研究电信诈骗犯罪手段发展规律

随着网络的发展和大数据的不断普及，电信诈骗的犯罪手段也在不断更新。采用文本分析方法可以发现电信诈骗犯罪手段的发展和变化规律：2016 年实施电信诈骗最常见的手段是进行电话诈骗。在电话诈骗中常见的术语是，冒用身份如受害人的熟人或是老友、“公检法”工作人员、银行工作人员、中奖信息客服、学校老师等，欺骗受害人把自己的财产打到诈骗分子的账户上去。此时，网络诈骗和短信诈

骗尚占少数。2017年诈骗手段更新换代, 诈骗分子搭建与银行网站极为相似的虚假网站, 通过群发网银密码器升级短信诱使受害人登陆假网站, 输入银行帐号、密码等信息, 犯罪分子在后台获取后, 再骗取动态口令, 迅速通过网银转账方式将受害人银行账户内资金转移。应用场景既包括冒充电商购物平台客服假装退款收取保证金的, 也包括直接通过广告短信、二维码等直接植入木马病毒, 盗取银行卡的帐号、密码等个人隐私信息, 最终实现以短信验证的方式诱骗得到手机验证码, 将受害者账户的资金转移走。2018年电信诈骗方式出现了两种明显的分化, 一种是兼职刷单诈骗, 诈骗分子允诺受害者帮助所谓的淘宝、京东、拼多多等网络购物平台的店铺完成“刷单”任务后奖励分红, 当受害人一开始收到回报后放松警惕, 渐渐敢于做一些涉及金额较大的“刷单任务”, 然后诈骗分子就会猝不及防地以操作失误等为借口, 骗取受害人用于刷单的“本金”。另一种则是针对老年人群的, 以网络推销医疗保健品为由, 以高价格哄骗受害者购买低成本的医疗保健品。到了2019年, 电信诈骗案件中最热门的高频词就是上文提到的“杀猪盘”, 诈骗分子利用网络社交软件和受害人培养感情, 诱导受害人参与网络赌博和使用网贷软件从而实施诈骗的一种行为。

纵观2016年以来的电信诈骗案件法案规律, 可以发现电信网络诈骗的发案趋势实现了从电话诈骗高发转变为网络诈骗高发的态势。电信诈骗实施的条件对热门网络平台比如购物平台、社交平台等的依赖性日趋增强。这为警方打击电信网络诈骗提供了渠道, 如可以通过与第三方平台的合作, 监测这些电信诈骗行为发生时惯用的“套话”等, 或是监测这些用于实施诈骗的账号、店铺商家等共同的特征从而实现打击新型的电信网络诈骗。在打击的同时, 也可以把这些电信诈骗行为发生时的“关键词”作好梳理归类, 用于对人们的防范电信诈骗的宣传教育, 警示人们当这些“关键词”“常用语”出现时, 面对自己的就是诈骗分子了, 争取从源头上提高人们的警惕性, 减少电信诈骗的发生, 避免人们财产的损失。通过文本分析也同样表明不同年份发生的电信诈骗案件, 有着不同的“关键词”, 这也就验证了通过文本分析的方式研究电子卷宗, 能够在统计不同类型电信诈骗案件的发案数量的基础上, 为公安机关研究电信诈骗案件发展规律和打击电信诈骗提供指导。

## 6. 结语

本文提供了一种对于研究电信诈骗案件的新的研究思路, 即以对于这类案件的电子卷宗进行文本分析, 归纳出其中的高频词汇, 通过筛选和联想揭示这一类型案件的特点以及发展规律, 并采用公安机关统计的数据进行参照对比。基于文本计量的案件研究新思路应用于公安机关案件规律研判具有良好效果。通过提炼高频词汇, 公安机关能够基本掌握此类案件的特点、常用词等, 在网络监测时, 一旦社交软件、平台上出现此类词汇, 就可以进行提前预警和检测。另外, 公安机关还能以此加强与网络社交平台的合作, 将这些“常用词”设置为敏感词汇, 当用户在平台上发出类似语句的时候, 可以被系统检测到并自动屏蔽, 可以一定程度上减少此类案件的发生。通过研究案件地区的发展规律, 沿海地区的公安部门更应加强警惕, 随时做好应对新型电信网络诈骗, 也可以遏制新型骗术的蔓延。这样一来, 本文的研究就能应用到公安部门打击电信诈骗案件的实战中去。

## 基金项目

江苏省高等学校大学生实践创新创业训练计划“公安案件电子卷宗文本计量研究——以电信诈骗案件为例”项目, 项目编号: 201910329042Y; 江苏省公安厅公安理论及软科学项目“公安案件电子卷宗文本大数据资源开发利用路径研究”, 项目编号: 2019LX013。

## 参考文献

- [1] 娄永涛, 唐祥. 大数据时代电信网络诈骗犯罪的防控反思[J]. 重庆理工大学学报, 2020(3): 121-128.

- [2] 王雪, 潘晓龙. 电信诈骗犯罪案件研究——以网络“杀猪盘”为例[J]. 法制博览, 2020(13): 17-21.
- [3] 许振亮, 刘喜美. 电信诈骗研究的关键词共词网络图谱分析[J]. 河南警察学院学报, 2017(6): 36-41.
- [4] 电信诈骗现新骗局 教你五大关键词锁定骗局[N/OL]. <http://www.ahwang.cn/china/20170526/1637800.shtml>, 2017-05-26.
- [5] 刘紫玟, 黄庆旭. 生态系统服务在土地利用规划研究和应用中的进展——基于文献计量和文本分析法[J]. 地理科学进展, 2019(2): 236-247.
- [6] 郭金子. 基于 CNKI 数据库的文献计量分析工具研究[J]. 图书馆学刊, 2014, 36(4): 113-117.
- [7] 杨晓宁, 黄丽娜. 我国大陆地区电信诈骗新特点及侦防对策[J]. 云南警官学院学报, 2016(4): 91-95.
- [8] 2016 年中国电信诈骗事件分析报告[N/OL]. [https://www.sohu.com/a/115556246\\_185924](https://www.sohu.com/a/115556246_185924), 2016-10-08.
- [9] 杨光, 沙贵君. 涉案类电信诈骗犯罪的特点及打防对策[J]. 河北公安警察职业学院学报, 2016, 16(3): 20-26.