

互联网信息服务分类方法研究

王宇航^{1,2}, 郭涛^{1,2}, 胡静远², 张潇丹^{1,2}, 韩冀中^{1,2}

¹中国科学院大学网络空间安全学院 北京 中国 100049

²中国科学院信息工程研究所 北京 中国 100093

摘要 互联网信息服务的蓬勃发展促进了互联网上信息内容的爆炸式增长和裂变式传播,同时也相应带来了安全挑战。对互联网信息服务进行分类管理是实现互联网信息内容安全针对性监管的有效手段,而目前国内外尚缺乏系统性的互联网信息服务分类方法。本文从互联网信息服务功能角度出发,提出一种基于信息生命周期的自顶向下划分的互联网信息服务分类体系,并从信息传播角度提出互联网信息服务分类方法。结合传播学相关理论并通过对我国主要互联网信息服务调研,验证了分类方法的合理有效性。同时,通过调研分析了我国互联网信息服务的信息传播功能特点,为互联网信息服务管理提供参考。

关键词 互联网信息服务; 分类方法; 信息内容安全; 服务功能; 信息传播

中图法分类号 TP309.2 DOI号 10.19363/J.cnki.cn10-1380/tn.2023.06.10

Classification Methods of Internet Information Services

WANG Yuhang^{1,2}, GUO Tao^{1,2}, HU Jingyuan², ZHANG Xiaodan^{1,2}, HAN Jizhong^{1,2}

¹School of Cyber Security, University of Chinese Academy of Sciences, Beijing 100049, China

²Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100093, China

Abstract The vigorous development of Internet information services has promoted the explosive growth and rapid dissemination of information content on the Internet, while also bringing corresponding security challenges. Classified management of Internet information services is an effective means to achieve targeted supervision of Internet information content security, but there is still a lack of systematic Internet information service classification methods at China and abroad. From the perspective of Internet information service functions, this study proposes a top-down classification system for Internet information services based on the information life cycle, and a classification method for Internet information services based on information dissemination. The effectiveness and rationality of the classification method is verified through the communication and investigation of major Internet information services in China. At the same time, we analyzed the functional characteristics of the dissemination information of China's Internet information services to provide references for Internet information service management.

Key words internet information service; classification method; information content security; service function; information dissemination

1 引言

近年来,随着互联网移动化、智能化发展,互联网终端结点不断增加,网络结构日趋庞大而复杂,互联网用户规模持续扩大,截至2020年1月,全球互联网用户超过45亿人,2019年增长7.75亿人,覆盖近60%人口,我国互联网用户规模超过9亿人,互联网普及率达64.5%^[1]。智能推荐、边缘计算、区块链、虚拟现实等新一代信息技术日新月异,促进了即时通信、网络直播、短视频等各类互联网信

息服务不断涌现,截至2019年底,我国互联网应用程序总量为367万款^[2]。互联网基于即时性、开放性、互动性等特点,日益成为信息产生和传播的主要渠道,全球社交媒体活跃用户数已超过38亿人,互联网用户人均每日在线时间长达6小时43分钟^[3]。互联网信息服务的推陈出新深刻改变了传统的信息单向传播、中心化传播方式,社交媒体的舆论影响力和社会动员力日渐增强,推动了信息爆炸式增长和裂变式传播。

互联网新技术、新应用的爆发式增长使得互联

通信作者: 郭涛, 博士生导师, 研究员, 工学博士, Email: guotao@iie.ac.cn。

本课题得到信息安全技术互联网信息安全服务通用要求(No. Y9V1301105, No. 90SNHTBH-2019111326), 2019年度互联网新技术新应用安全评估-1项目(No. Y9V1501105)资助。

收稿日期: 2020-11-02; 修改日期: 2020-12-31; 定稿日期: 2025-11-28

网信息服务在信息内容安全性方面的考虑往往并不完备,加之部分互联网信息服务提供者在利用这些新技术、新应用提供信息服务时的不负责任和别有用心,信息内容安全正在成为互联网信息安全风险的主要因素之一。仅 2020 年 4 月,全国各级网络举报部门受理网络违法和不良信息举报高达 1458.1 万件,同比增长 20.9%^[4]。

我国在互联网信息服务管理方面的实践处于起步阶段,长期以来主要采取对提供违法有害信息的互联网信息服务进行处置的“事后”监管模式,此种监管模式不可避免地造成不良后果“先产生、后治理”的被动局面,探索在违法有害信息发布传播前的“事前”监管模式迫在眉睫。“事后”监管实现相对容易,主要依据相关法律法规判断违法违规情况并进行处置即可,而“事前”监管则需要对潜在信息安全风险进行审慎预判。

互联网信息服务种类繁多,功能聚合特性日趋明显,不同类型服务功能的信息传播媒介功能不同,传递信息的类型不一,面向的用户群体和范围不同,产生的传播效果和舆论动员能力迥异,由此带来的信息内容安全风险程度不一,需要对不同类型的互联网信息服务进行分类管理。而目前国内在互联网信息服务分类方面实践研究不足,尚缺乏从信息传播和信息内容安全角度的互联网信息服务系统性分类方法,各互联网相关统计管理机构和应用商店等在对互联网信息服务按功能分类时往往根据其最主要功能特征进行简单归类,分类尺度不等、分类标准不一、分类体系不合理。国内外面向互联网信息内容安全的分类研究仅仅局限于针对信息内容本身^[5-10],如国际上美国 RSACi 网络分级系统、电影内容分级,英国的网络内容选择平台(PICS),德国的《青少年媒介保护国家条约》,我国台湾地区的《电脑网路内容分级处理办法》,我国《绿色游戏推荐标准》等。

本文第 2 节首先探讨了互联网信息服务的概念,并对国内互联网信息服务分类现状进行阐述;第 3 节从互联网信息服务功能角度出发,提出一种基于信息生命周期的自顶向下划分的互联网信息服务功能分类体系,并提出基于信息传播功能的互联网信息服务分类方法;第 4 节通过对国内 792 款主要互联网信息服务开展调研,验证了互联网信息服务分类体系的有效性和合理性,依据该分类方法对上述互联网信息服务进行分类,并对互联网信息服务功能及服务形式的特点进行分析探讨;第 5 节对全文进行总结并提出下一步工作计划。

2 互联网信息服务分类现状

2.1 互联网信息服务的概念

在信息化时代,信息和信息技术极大地改变了人们的生产生活,信息已经成为一种具有战略意义的基础资源。信息是事物的一种属性^[11],信息的表达需要借助文本、图形、音频、视频等约定的符号,信息的存储和传播需要借助纸张、磁盘、光盘、网络线路等介质。数字信息的载体可划分为三个层次:信息技术或产品、信息系统或网络、信息服务^[12],信息技术或产品是指路由器、操作系统、数据库等存储和使用信息数据的基本模块;信息系统或网络是指存储和使用信息数据的全体基础设施和组织管理的总和^[13];信息服务是面向用户提供信息内容的服务,如文字处理软件 Office、多媒体播放器 QQ 音乐、网页浏览器 IE 等。《互联网信息服务管理办法》^[14]对互联网信息服务的定义为“通过互联网向上网用户提供信息的服务活动”。在互联网中,各类信息服务直面互联网用户,提供信息采集发布、使用传播、存储销毁等主要服务,是互联网信息生态的核心组成部分。

2.2 互联网信息服务分类现状

互联网信息服务种类繁多,在终端形态上的演变主要经历了从 20 世纪 90 年代互联网站的兴起,到 2010 年前后移动应用程序的蓬勃发展,再到 2014 年起轻应用的大量涌现 3 个阶段^[15]。目前,在行业层面,数据统计机构和各类应用商城对互联网信息服务进行的分类均是从自身业务角度出发,关注角度如主要服务群体、提供信息类型、特色服务功能等。表 1 列举了部分数据统计机构和各类应用商城对互联网信息服务的分类,可以看到分类尺度不等、维度不一、范围重叠、名称混用等是目前互联网信息服务分类的主要现象。此外,从保障互联网信息内容安全角度,目前行业内也缺乏对可能带来信息内容安全风险的信息服务功能的定义和分类等相关研究。如何对互联网信息服务进行合理分类管理是一项亟须解决的问题。

在国家管理层面,2020 年 1 月,全国信息安全标准化技术委员会发布国家标准《信息安全技术 互联网信息服务安全通用要求》征求意见稿^[21],借鉴了《信息技术安全评估准则》(GB/T 18336—2015)^[22]的组件化描述方式,首次面向互联网信息服务提出保障信息内容安全应满足的通用要求。该标准在附录 B 中以组件包形式列出了互联网信息服务的信息发布等服务形式,突破以往行业内基于服务主营业务的主观式分类分析方法,首次从服务功能形式级别探索互联网信息服务的分类方法。

表 1 互联网信息服务分类举例

Table 1 Examples of Internet information service classification

来源	《互联网新闻信息服务管理规定》 ^[16] 对信息服务的分类	艾瑞数据 ^[17] 对互联网站的分类	CNNIC ^[1] 对移动应用程序的分类	Quest Mobile ^[18] 对微信小程序的分类	华为移动应用商城 ^[19] 的分类	小米移动应用商城 ^[20] 的分类
类目	互联网站	综合资讯	游戏	新闻资讯	购物比价	游戏
	应用程序	电子商务	日常工具	实用工具	影音娱乐	实用工具
	论坛	社交网络	电子商务	健康美容	实用工具	影音娱乐
	博客	实用工具	生活服务	出行服务	便捷生活	聊天社交
						

3 互联网信息服务分类方法

本节从互联网信息服务信息内容安全分类管理的角度,探讨基于信息生命周期的自顶向下划分的互联网信息服务分类方法。3.1 节探讨了互联网信息服务在信息各生命阶段提供的功能,提出互联网信息服务分类模型——莲藕模型;3.2 节从传播信息这一最重要的信息生命阶段出发,提出互联网信息服务信息传播功能划分方式,并借助传播学相关理论分析验证该分类方法和分类结果的合理有效性;3.3 节基于上述信息服务功能的分类方式,提出互联网信息服务的一般性分类方法。

3.1 互联网信息服务功能分类模型

互联网信息服务是一种向互联网用户提供信息的服务活动,由若干支撑信息活动的服务功能组成。文献[12]将信息生命周期划分为产生、处理、发布、传播、存储和销毁 6 个阶段,相应地,互联网信息服务功能也可按支撑信息活动的各阶段划分为产生信息、处理信息、发布信息、传播信息、存储信息和销毁信息 6 类功能,一款互联网信息服务应由上述 6

类功能中的若干功能组成。其中某 1 类功能,可以根据不同需要从不同角度进行细分,如产生信息类功能可以按产生信息的用户主体类型分为面向人类用户的功能和面向信息技术实体的功能(生成信息功能),其中面向人类用户的功能可进一步按产生信息的形式分为非文件形式(填写信息功能)和文件形式(上传文件功能)等;又如传播信息类功能可分为话题、通信、群组、直播等具体的功能。

本文提出互联网信息服务功能结构“莲藕模型”(图 1),用以形象地表示出互联网信息服务提供信息服务的活动过程。莲藕模型将互联网信息服务抽象为一段莲藕链,包含互联网信息服务的全体功能,莲藕链包含 6 节莲藕,每一节莲藕表示按信息生命阶段划分的 1 类服务功能;沿着莲藕链从 A 至 B 的方向表示信息的流动方向,每节莲藕垂直于莲藕链方向的切面包含数量可能不等的若干小孔,表示包含在该类功能中的功能,每一节各个小孔在节端点处汇聚成一个大孔,相邻两节莲藕间联通;信息服务提供的信息按生命周期从莲藕链的 A 端开始,流动穿过每节莲藕的某些小孔,最终停止于莲藕链的 B 端。

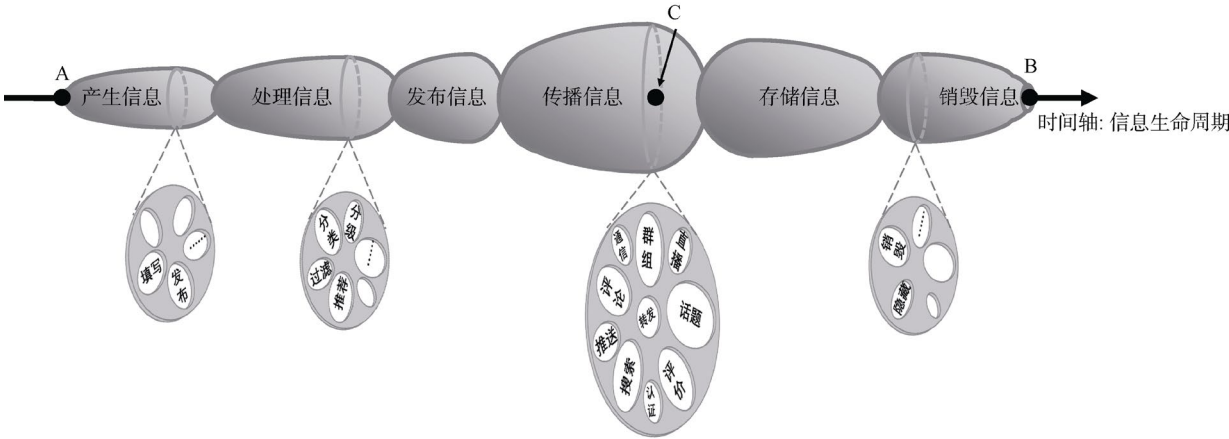


图 1 互联网信息服务功能分类模型——莲藕模型

Figure 1 A lotus-root-like model for classification of Internet information service functions

互联网信息服务主要相关方有服务提供者、服务使用者和服务监管者等,各相关方对不同服务信息生命阶段的关注程度有所不同。从服务提供者视角看,为保障互联网信息服务安全有效运营,服务

提供者需对服务信息全生命周期予以足够关注和支持;从服务使用者视角看,用户主要关注和使用服务信息产生、发布、传播和销毁阶段的服务功能;从服务监管者视角看,为保障互联网信息内容安全,同样需从信息全生命周期进行监督管理,如国家标准《信息安全技术 互联网信息服务安全通用要求》征求意见稿中对 6 个信息生命阶段的服务功能均设置了相应安全要求组件。

3.2 互联网信息服务信息传播功能分类方法

根据莲藕模型(图 1),互联网信息服务提供的信息从 A 点开始,经过产生、处理、发布和传播 4 个阶段,于 C 点完成信息的大规模传播,若此信息包含违法不良内容,则将在 C 点发生信息安全事件,产生不良后果,“事前”监管应在信息由 A 点到 C 点的行进过程中采取管理措施,避免违法不良信息在信息生命周期时间轴上行进到 C 点。从信息生命周期的各环节角度看,“事前”监管在信息产生、处理、发布和传播 4 个环节的任意 1 个环节采取有效管理措施,即可在“事前”遏制违法不良信息的传播。

社会动员是人类社会特有的能动性活动^[23],互联网信息服务的社会动员能力主要取决于其传播信息的能力,互联网信息服务主要相关方也均对其信息传播功能尤为关注。为实现对互联网信息服务信息内容分发的精细化分类管理,本文选择从信息传播角度出发,研究互联网信息服务的信息传播功能的分类方法,以利于在不同传播形式下对信息内容实施不同程度的监测识别和过滤管理。

3.2.1 分类方法

树是一种不含回路的图结构,决策树^[24]是一种树形结构,是一种常用的分类方法。决策树的每个内部结点进行关于某属性的划分操作,每个分支对应于该划分操作的一个子集,通过自顶向下的决策方式,将每个结点划分为若干子结点,从而将总体进行分类。设结点 b 为结点 a 的子结点, U_a 、 U_b 分别表示结点 a 、 b 的全部样本集合, p 表示样本关于某属性的一个划分条件, $p(x)$ 表示样本 x 满足条件 p , 定义划分操作 f , 使得

$$U_b = f(U_a, p) = \{x | p(x), x \in U_a\} \quad (1)$$

麦肯锡咨询公司高级顾问 Minto^[25]提出 MECE (Mutually Exclusive Collectively Exhaustive) 分类原则,指出分类应确保每个类别“相互独立”,并对总体“完全穷尽”。为实现 MECE 原则的“相互独立”和“完全穷尽”目标,定义一组互补划分条件($p_1, p_2, \dots, p_c$),使得

$$\begin{cases} \forall i \in (1, 2, \dots, c), j \in (1, 2, \dots, c), i \neq j, \\ U_{a_i} \cap U_{a_j} = \emptyset \\ \forall i \in (1, 2, \dots, c), \bigcup U_{a_i} = U_a \end{cases} \quad (2)$$

其中, $U_{a_i} = f(U_a, p_i)$ 。分类决策树的构造以样本总体为根结点,以自顶向下的顺序,对每个结点设置一组互补划分条件并实施划分操作,每一次划分的结果均满足 MECE 原则。

3.2.2 信息传播功能分类

本文提出互联网信息服务信息传播功能分类模型(图 2),以信息的传播属性为划分条件,参考物理学上简谐波的定义要素(波源、方向、振幅、频率),以功能使用者为信息传播的参照点,按照传播方向、传播范围和传播频次 3 个条件,将互联网信息服务提供的信息传播功能划分为 4 个类目。为保障划分结果满足 MECE 原则,统一采用“二分法”,每次划分通过设置一组互补划分条件(如 p 和 $\neg p$)实施划分操作。按照传播方向、传播范围和传播频次 3 个传播属性,定义条件 p_1 、 p_2 、 p_3 (表 2),将互联网信息服务的信息传播功能(U_p)依次按上述条件及其互补条件进行划分,分别得到消息(U_1)、直播(U_2)、通信(U_3)和搜索(U_4) 4 类功能,具体划分计算公式和类别描述见表 3。在传播学中,传播者主动传播信息的方式一般按照传播范围分为自我传播、人际传播、群体传播和大众传播 4 种^[26],其中自我传播是个体对信息的加工过程,不在互联网信息传播的范围。其余 3 种社会传播类型中,人际传播和群体传播分别指 2 个

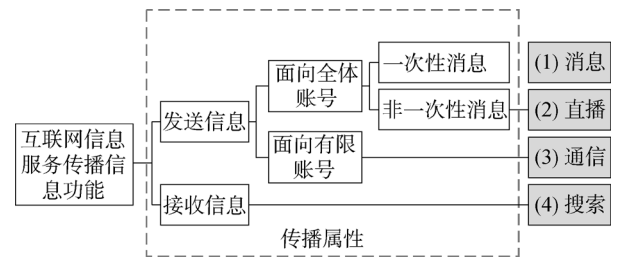


图 2 信息传播功能分类模型(4 类)

Figure 2 Classification model of information dissemination functions (4 types)

表 2 基于传播属性的划分条件定义

Table 2 Definition of classification conditions based on dissemination attribute

条件	描述
p_1	功能是功能使用者用以发送信息的
p_2	功能是面向全体账号的
p_3	功能是一次性传播信息的

表 3 信息传播功能(U_p)划分结果

Table 3 Classification results of Information dissemination functions (U_p)				
序号	类别名称	划分表达式	表达式含义	一般理解
1	消息	$U_1=f\left(f\left(f\left(U_p,p_1\right),p_2\right),p_3\right)$	面向全体公众账号的一次性发送信息功能	面向公众账号发布信息的功能(如博文发布、转发、评论、评价、弹幕等)
2	直播	$U_2=f\left(f\left(f\left(U_p,p_1\right),p_2\right),\neg p_3\right)$	面向全体公众账号的非一次性发送信息功能	面向公众账号持续发布信息的功能(如广播、视频直播等)
3	通信	$U_3=f\left(f\left(U_p,p_1\right),\neg p_2\right)$	非面向全体公众账号的发送信息功能	面向有限账号发送信息的功能(如发短信、通话、群聊等)
4	搜索	$U_4=f\left(U_p,\neg p_1\right)$	非发送信息的信息传播功能	索取信息的功能(如搜索、订阅等)

及以上个体和群体组织个体间的传播活动,在以互联网为媒介的传播方式中可共同归纳为通信(U_3);大众传播是指面向大众受传者的传播活动,在互联网传播方式中由消息(U_1)和直播(U_2)组成。受传者主动索取信息的活动统一归纳为搜索(U_4)。

进一步地,可将消息类功能(U_1)按照传播消息的增益性、独立性 2 个消息属性,继续划分为 3 类(图 3)。定义条件 q_1 、 q_2 (表 4),将消息类功能(U_1)依次按上述条件及其互补条件进行划分,分别得到话题(U_{1-1})、评议(U_{1-2})、转发(U_{1-3}) 3 类功能,具体划分计算公式

和类别描述见表 5。在这一层面上,互联网信息服务信息传播功能被划分为满足 MEMC 原则的话题(U_{1-1})、评议(U_{1-2})、转发(U_{1-3})、直播(U_2)、通信(U_3)和搜索(U_4) 6 类功能。国家标准《信息安全技术 互联网信息服务安全通用要求》征求意见稿^[22]的附录 B 中列举的“常见互联网信息服务形式的安全技术要求组件包”中,信息发布、评论评价、信息分享、网络直播、通信群组和内容搜索 6 个组件包与互联网信息服务信息传播功能在这一层面的 6 类划分是一致的。

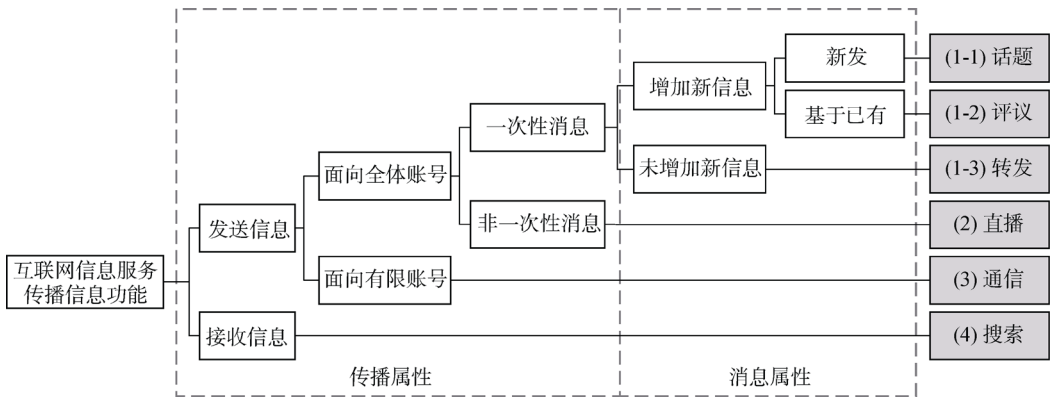


图 3 信息传播功能分类模型(6 类)

Figure 3 Classification model of information dissemination functions (6 types)

表 4 基于消息属性的划分条件定义

Table 4 Definition of classification conditions based on message attribute	
条件	描述
q_1	消息增加了新信息
q_2	消息不依赖于其他消息

更进一步地,评议类功能(U_{1-2})可按信息属性“是否包含信息内容”划分为包含信息内容的评论类功能(如留言、弹幕)和不包含信息内容的评价类功能(如点赞、打赏),评价类功能还可进一步按内容属性“权威性”划分为权威性评价(如认证)和非权威性评价(一般性评价)等。互联网信息服务分类方法可良好地满

足任意精度的划分方式,本文不再进行深入讨论。

3.2.3 几种常见信息传播功能分析

“媒介就是插入传播过程中,用以扩大并延伸信息传送的工具”^[27],信息服务的功能可以看作一种媒介。传播学之父 Schramm 等^[27]提出了认识媒介的 8 个原则,即刺激感官数量、反馈机会、传播速度控制、信息具象化能力、信息增殖能力、信息保存能力、克服弃取能力和满足专门需要能力(表 6)。研究^[27-28]列举了 23 种常见信息传播功能(表 7),分别按照上述 8 个媒介属性,依据表 6 所列量化标准进行量化评分,结果见表 8。

本文借助经典聚类算法 K-means 的改进算法 K-means++^[29](优化了初始类中心的设置),将上述 23

表 5 消息类功能(U_1)划分结果

Table 5 Classification result of message functions (U_1)

序号	类别名称	划分表达式	表达式含义	一般理解
1	话题	$U_{1-1}=f(f(U_1,q_1),q_2)$	面向全体公众账号一次性新发增加新信息 的消息	面向全体公众账号发布独立消息的功能(如 博文发布、话题创建等)
2	评议	$U_{1-2}=f(f(U_1,q_1),\neg q_2)$	基于已有信息面向全体公众账号一次性发 送增加新信息的信息	针对已有信息发布信息的功能(如评论、弹 幕、点赞、认证、竞价排名等)
3	转发	$U_{1-3}=f(U_1,\neg q_1)$	面向全体公众账号一次性发送未增加新信 息的消息	转发已有信息的功能(如转发、点名等)

表 6 认识媒介的原则

Table 6 Principles for understanding media

序号	媒介属性	描述	量化标准
1	刺激感官数量	信息符号刺激受众感官方 式的种类数量	依据可包含听觉、图文视觉、视频视觉 3 种感受途径情况, 包含 1 种计 0 分, 2 种 计 0.5 分, 3 种计 1 分
2	反馈机会	受传者可作出反馈的机会 大小	依据传播者可接收受传者可反馈性及反馈速度, 0 分表示不可收到反馈, 0.5 分表示 可延迟收到反馈, 1 分表示可实时收到反馈
3	传播速度控制	受传者可控制信息获取速 度的能力	依据受众可自主获取信息的速度, 0 分表示信息传播速度由受传者主导, 0.5 分表示 信息传播速度由传播者和受传者共同控制, 1 分表示信息传播速度由传播者主导
4	信息具象化能力	媒介允许的信息具象化表 达的能力	依据媒介主要包含信息符号情况, 0 分表示仅含信号或文字符号, 0.5 分表示包含图 像或音频符号, 但不含全景图或视频符号, 1 分表示包含全景图或视频符号
5	信息增殖能力	媒介允许的信息可分享的 能力	依据媒介中信息分享的代价, 0 分表示人际传播, 0.5 分表示群体传播, 1 分表示大 众传播
6	信息保存能力	媒介保存信息的能力	依据媒介保存信息的能力, 0 分表示信息符号转瞬即逝, 0.5 分表示信息符号可在 一定时间内或一定条件下保存, 1 分表示信息符号可几乎永久保存
7	克服弃取能力	媒介克服受众拒绝接受传 播信息的能力	依据受众拒绝接受或中断传播信息渠道的代价, 0 分表示低代价(如关闭网页等随 意弃取行为), 0.5 分表示中等代价(如退出群聊等弃取可被传播者察觉的行为), 1 分 表示高代价(如拒绝电话等中断传播的行为)
8	满足专门需要能力	媒介针对不同受传者传播 特定内容信息的能力	依据媒介允许传播者针对不同受传者定向传播信息的能力, 0 分表示传播者无法 定向传播信息, 0.5 分表示传播者可选择一类可能的受众群体, 1 分表示传播者可针 对特定受传者传播信息

表 7 常见信息传播功能

Table 7 Common information dissemination functions

序号	功能名称	说明或示例
1	点发消息	点对点发送文字、图片、语音消息, 如点对点即时通信、点对点邮件发送
2	群发消息	群发文字、图片、语音消息, 如群发短信、邮件抄送
3	点名推送	示例: 新浪微博“@”
4	语音通话	网络点对点语音通话
5	语音群聊	多人语音会议
6	视频通话	网络点对点视频通话
7	视频群聊	多人视频会议
8	位置共享	双人或多人位置信息实时共享, 示例: 微信“共享实时位置”
9	图文直播	示例: 百度贴吧“文字直播”
10	音频广播	网络广播电台
11	视频直播	网络视频直播
12	博文发布	示例: 微信朋友圈, 新闻客户端发布新闻, 知识问答平台提问
13	话题创建	示例: 新浪微博“话题”
14	转载发布	服务内外的消息、用户信息的转载发布
15	发起投票	示例: 新浪微博“投票”, 各类微信投票小程序
16	留言评论	跟帖评论、账号留言等
17	视频弹幕	视频分发平台的弹幕功能, 示例: 爱奇艺“弹幕”

续表

序号	功能名称	说明或示例
18	直播弹幕	视频直播平台的弹幕功能, 示例: YY 直播“弹幕”
19	点赞打赏	受众一般性评价行为, 示例: 哔哩哔哩“点赞”、“投币”, 简书“打赏”
20	权威认证	平台对信息内容或用户账号的权威认证, 示例: 百度搜索“官方网站”、百度百科“权威专家认证词条”, 新浪微博“认证用户”
21	竞价排名	搜索引擎竞价排名、广告投放排名, 示例: 新浪微博“推广”
22	消息搜索	服务内外的消息、用户、网络资源搜索
23	订阅关注	对博文、话题订阅收藏, 对用户添加好友、关注

表 8 常见信息传播功能的媒介功能评分

Table 8 Media functions score of common information dissemination functions

功能名称	媒介功能							满足专门需要能力
	刺激感官数量	反馈机会	传播速度控制	信息具象化能力	信息增殖能力	信息保存能力	克服弃取能力	
点发消息	0.5	0.5	0.5	0.5	0	0.5	0.5	1
群发消息	0.5	0.5	0.5	0.5	0	0.5	0.5	1
点名推送	0	1	0.5	0	0	1	0.5	1
语音通话	0	1	0.5	0.5	0	0	1	1
语音群聊	0	1	1	0.5	0.5	0	0.5	1
视频通话	0.5	1	0.5	1	0	0	1	1
视频群聊	0.5	1	1	1	0.5	0	0.5	1
位置共享	0	1	1	0.5	0.5	0	0.5	1
图文直播	0	0.5	0.5	0.5	1	0.5	0	0.5
音频广播	0	0.5	1	0.5	1	0	0	0.5
视频直播	0.5	0.5	1	1	1	0	0	0.5
博文发布	1	0.5	0	0.5	1	1	0	0.5
话题创建	0.5	0.5	0	0.5	1	1	0	0.5
转载发布	1	0.5	0	0.5	1	1	0	0.5
发起投票	1	0.5	0	0.5	1	1	0	0.5
留言评论	0.5	0.5	0	0.5	1	1	0	0.5
视频弹幕	0	0	0	0	1	1	0	0.5
直播弹幕	0	0	0	0	1	0.5	0	0.5
点赞打赏	0	0	0	0	1	1	0	0.5
权威认证	0	0	0	0	1	1	0	0.5
竞价排名	0	0	0	0	1	1	0	0.5
消息搜索	0.5	0	0	0	0	1	0	0
订阅关注	0	0	0	0	0	1	0	0

种信息传播功能聚为 4 类, 聚类结果见表 9。从各类包含的信息传播功能来看, 类 1 至类 4 分别对应通信(U_3)、直播(U_2)、消息(U_1)和搜索(U_4)。t-SNE^[30]是一种非线性降维机器学习算法, 常用于高维数据的可视化。借助 t-SNE 算法将上述信息传播功能的媒介属性评分降至 2 维, 并用不同颜色区分 4 类聚类结果绘制散点图(图 4)。由图 4 可见, 借助传播学媒介属性分析, 所举常见信息传播功能按照 4 类功能自然聚类。基于传播学角度的常见信息传播功能聚类结果与基于互联网信息服务信息传播功能分类模型(图 2)

所得分类结果一致, 从一定程度上验证了该分类方法的合理有效性。

3.3 互联网信息服务分类

本节从信息传播功能角度出发, 探讨互联网信息服务分类。按照互联网信息服务信息传播功能分类模型(图 2), 互联网信息服务的信息传播功能(U_p)被分为消息(U_1)、直播(U_2)、通信(U_3)和搜索(U_4) 4 类。在互联网信息技术的不断成熟和网络用户对互联网信息服务个性化需求不断增加的今天, 互联网信息服务发展迅速, 功能数量不断增加, 功能种类

表 9 常见信息传播功能聚类结果

Table 9 Clustering results of common information dissemination function				
类别序号	1	2	3	4
类成员	点发消息 群发消息 点名推送 语音通话 语音群聊 视频通话 视频群聊 位置共享	图文直播 音频广播 视频直播	博文发布 话题创建 转载发布 发起投票 留言评论 视频弹幕 直播弹幕 点赞打赏 权威认证 竞价排名	消息搜索 订阅关注
刺激感官数量	0.25	0.17	0.40	0.25
反馈机会	0.88	0.50	0.25	0
传播速度控制	0.69	0.83	0.00	0
信息具象化能力	0.56	0.67	0.25	0
信息增殖能力	0.19	1	1	0
信息保存能力	0.25	0.17	0.95	1
克服弃取能力	0.63	0	0	0
满足专门需要能力	1	0.50	0.50	0

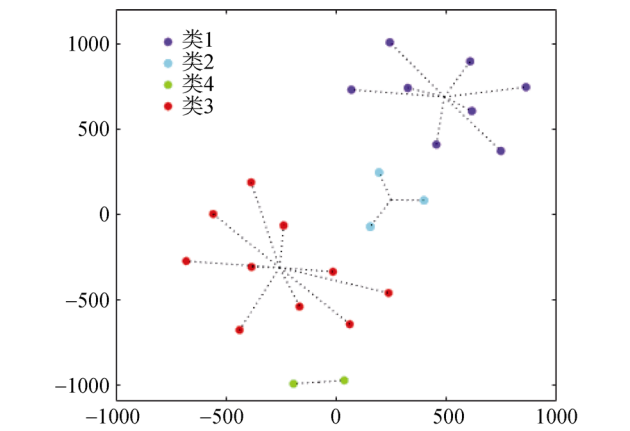


图 4 基于 t-SNE 降维表示的常见信息传播功能分布情况

Figure 4 Common information transmission function based on expression of t-SNE dimension reduction

呈现聚合、多样特征,从开展服务功能的角
度对互联网信息服务进行分类不能简单将其
归为某一类。本文提出以多标签描述方式
对互联网信息服务的类别进行描述,将互联
网信息服务按照是否包含上述 4 类信息传
播功能进行标记,此种方式对互联网信息
服务分类情况描述较为清晰,实际归类时
可操作性较强。表 10 列举了若干互联网
信息服务的分类情况。

同时,依据其包含的类别标签数量,可进
一步从互联网信息服务功能聚合程度角度
分为强综合功能型、弱综合功能型、单一
功能型 3 类,划分依据为

$$T(x)=\begin{cases} \text{强综合功能型}, & x>3 \\ \text{弱综合功能型}, & 2\leq x\leq 3 \\ \text{单一功能型}, & x<2 \end{cases} \quad (3)$$

其中, x 表示标签数量。

表 10 常见互联网信息服务分类情况

Table 10 Common types of Internet information services

类型	服务形态	服务名称	信息传播功能			
			消息	通信	直播	搜索
强聚合功能型	互联网站	淘宝网-网络购物	1	1	1	1
	移动应用程序	抖音短视频	1	1	1	1
	轻应用	拼多多	1	1	1	1
	互联网站	豆瓣-社区交友	1	1	0	1
弱聚合功能性	移动应用程序	微信	1	1	0	1
	轻应用	斗鱼直播	1	0	1	1
	互联网站	腾讯-在线视频	1	0	0	1
	移动应用程序	腾讯欢乐麻将	1	1	0	0
单一功能型	轻应用	智行火车票	1	0	0	1
	互联网站	中华网-新闻资讯	1	0	0	0
	移动应用程序	360 手机卫士	1	0	0	0
	轻应用	微信发票助手	0	0	0	1

4 互联网信息服务类别分析

本节选取 792 款国内主要互联网信息服务作为研究对象,通过人工逐一查验方式统计各服务包含的各类功能,分析互联网信息服务功能特点以及分类情况。

4.1 研究对象

互联网站、移动应用程序和轻应用是目前互联网信息服务的 3 类主要形态^[15]。本文按均匀比例原则,从上述 3 类服务形态中选取社会影响力排名靠前的 792 款信息服务作为研究对象,其中互联网站类 320 个、移动应用程序类 213 个、轻应用类 259 个,图 5(a)显示了选取上述 3 类服务形态的样本占比情况。

互联网站方面,根据中国互联网络信息中心(CNNIC)2020 年 4 月发布的第 45 次《中国互联网络发展状况统计报告》^[1],截至 2020 年 3 月我国网民总数为 9.04 亿人。根据知名移动互联网第三方数据机构艾瑞咨询公司 2020 年 3 月统计数据^[17],选取覆

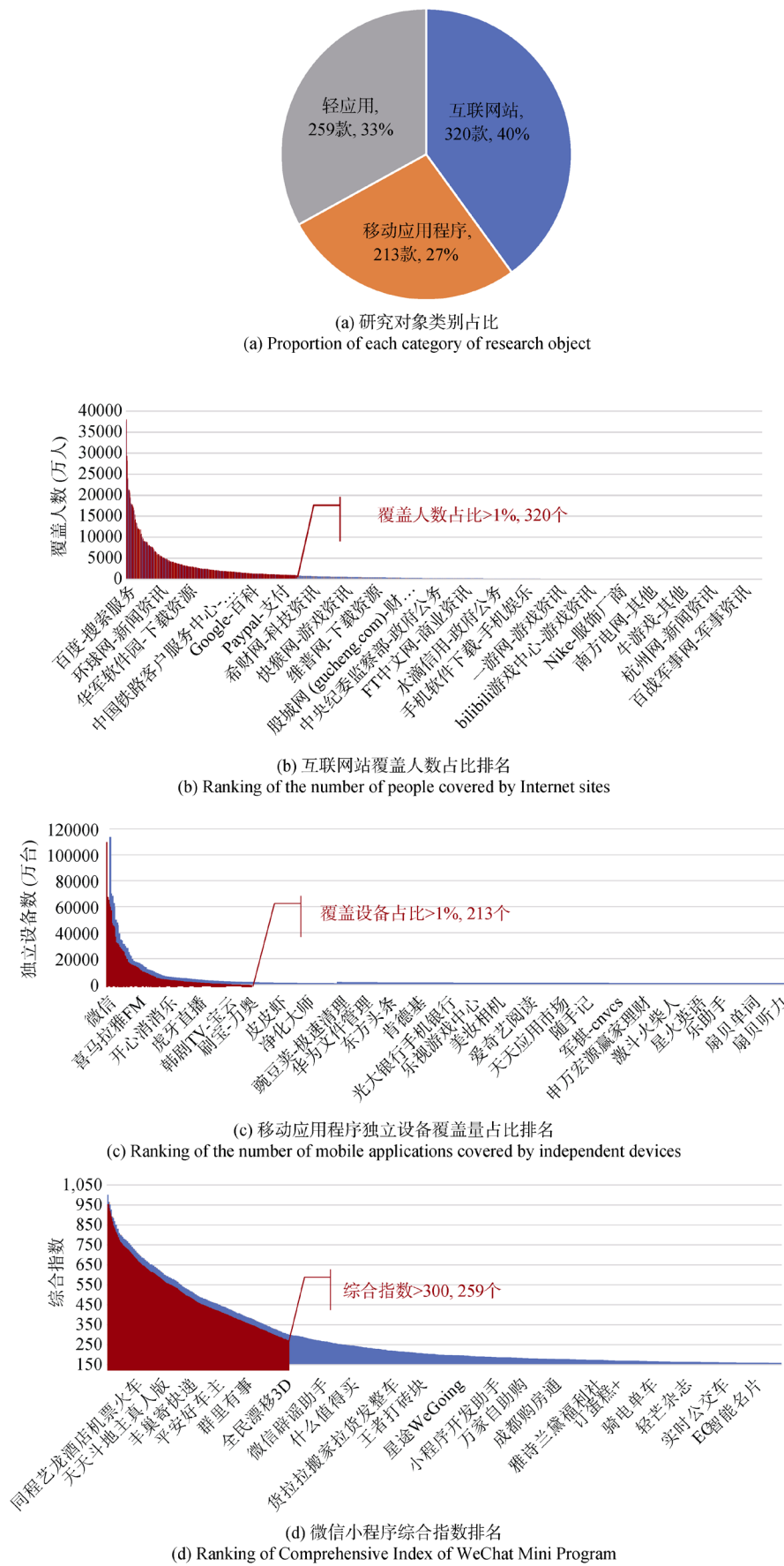


图 5 研究对象选取情况

Figure 5 Research objects

盖人数占我国网民总数比例排名前 1% 以上的 320 个互联网站。图 5(b) 显示了互联网站覆盖人数占比情况, 可见互联网站覆盖人数具有长尾效应, 其中红色部分为占比超过 1% 的互联网站。

移动应用程序方面, 根据人民网 2020 年 7 月发布的《中国移动互联网发展报告(2020)》^[2], 2019 年我国移动设备总数为 15.7 亿台。根据艾瑞咨询公司 2020 年 3 月统计数据^[28], 选择覆盖设备占我国移动设备总数比例排名前 1% 以上的 213 个移动应用程序。图 5(c) 显示了移动应用程序覆盖设备数占比情况, 同样具有长尾效应, 其中红色部分为占比超过 1% 的移动应用程序。

轻应用方面, 根据 2019 年 8 月国内移动互联网大数据公司 QuestMobile 发布的《小程序 2019 半年报告》^[18], 截至 2019 年 6 月国内已有微信、百度、支付宝等 8 家移动互联网平台推出小程序, 其中微信小程序具有先发优势且最成熟, 总体用户量达 7 亿人。为排除不同平台对统计结果的差异化影响, 我们选择用户规模最大的微信平台作为统一尺度, 以微信小程序作为轻应用的研究对象, 根据艾瑞咨询公司统计数据^[31], 经综合评价活跃设备数、成长趋势、稳定性等多个指标, 选取评分指数高于 300 的微信小程序 259 个。图 5(d) 显示了上述指数情况, 其中红色部分为指数高于 300 的微信小程序。

4.2 信息传播功能分析

本节通过人工逐一查验上述 792 款互联网信息服务提供的功能, 依据表 5 划分的 6 类信息传播功能定义, 将提供相应功能的信息服务进行统计。调研未发现无法归类的功能, 验证了互联网信息服务功能分类模型的合理性。

图 6 统计了 6 类信息传播功能在 3 类互联网信息服务形态中的占比情况, 并按平均占比降序排列。占比超过 80% 的功能是话题和搜索 2 类, 分别为互联网用户发送和接收信息的最主要形式; 占比超过 30% 不足 80% 的功能是转发和评议, 是两类二次传播信息的功能; 占比小于 30% 的最后 2 类分别是通信和直播, 这些是具有特殊传播性质的功能。

从图 6 中还可以看到, 话题和搜索 2 类基本信息传播功能在互联网网站形态的服务中占比最高, 其余功能在移动应用程序形态的服务中占比最高, 而各类功能在轻应用形态的服务中占比均最低, 可见互联网站更加侧重信息互联基础服务, 移动应用程序提供更加多样化的信息传播功能, 而轻应用因其自身的轻量级定位, 往往侧重于提供少量专门性功能。

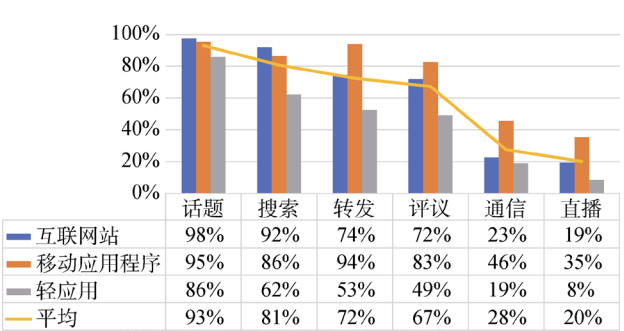


图 6 互联网信息服务的 6 类信息传播功能统计
Figure 6 Six information dissemination function types of Internet information services

4.3 服务分类分析

4.3.1 服务分类分布情况

图 7 统计了 792 款互联网信息服务的分类情况, 按照消息、通信、直播和搜索 4 类服务形式在互联网网站、移动应用程序和轻应用 3 种服务形态的服务中平均占比降序排列。从平均占比来看, 消息和搜索 2 类互联网用户发送和接收信息的最主要服务形式占比最高, 分别为 97.7% 和 80.4%, 通信和直播 2 类特殊传播信息服务形式占比较低, 分别为 27.4% 和 20.0%。从 3 类互联网信息服务形态来看, 消息服务形式在 3 类服务形态的服务中占比均较高, 搜索服务形式在互联网网站中占比最高, 通信和直播服务形式在移动应用程序中占比均最高, 可以认为互联网站形态的服务更加侧重信息聚合, 而移动应用程序形态的服务更加侧重用户互动。

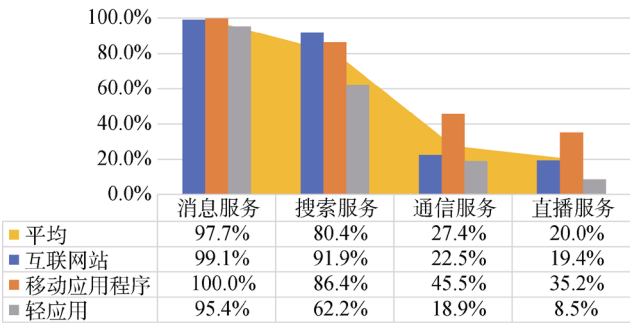


图 7 互联网信息服务的 4 类服务形式统计
Figure 7 Four service forms of Internet information services

图 8 探讨了 792 款互联网信息服务的消息、通信、直播和搜索 4 类服务形式组合分布情况。互联网信息服务是否包含 4 类服务形式的组合有 16 种可能, 为直观表达各类服务形式的组合分布情况, 图 8 借助二进制编码思路, 使用 4 位二进制数表示这 16 种情况, 按照消息、通信、直播和搜索的顺序由高位

起逐位编码, 每一位数字为 1 表示该服务包含此类服务形式, 否则为 0; 编码按照红、橙、黄、绿和蓝色的顺序依次表示包含 4 至 0 种服务形式的情况。借助集合 Venn 图思想, 用 4 个实线圆形表示上述 4 类服务形式, 通过圆形两两相交方式将平面划分为 14 个部分, 每个部分分别代表一种组合方式; 由于在二维平面两两相交的 4 个圆形中互为对角的 2 个圆形无法单独重叠, 因此借助一个虚线圆形将 4 个集合的独自区域划分为 2 部分, 虚线圆形内部区域表示互为对角圆形代表集合的相交情形, 虚线圆形外部区域表示 4 个圆形各自区域的单独情形; 为保持图形整体呈中心对称以使图形整体从中间到周围的集合重合度依次降低, 方便观察 4 类服务形式的组合分布情况, 本文没有采用更为严格但图形不规则的其他 4 集合 Venn 图方案。为方便描述, 将图形放置在二维直角坐标系中, 4 个集合依次按照第一象限至第四象限的顺序表示消息、通信、直播和搜索服务形式。

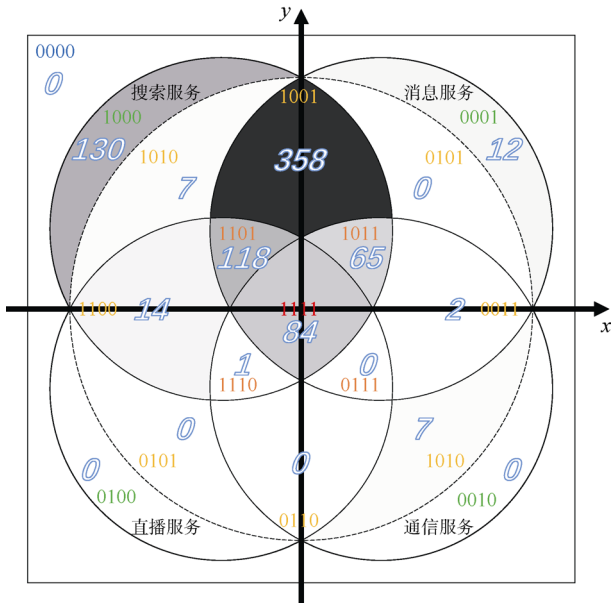


图 8 互联网信息服务的 4 类服务形式组合分布情况
Figure 8 Distribution of combinations of four service forms of Internet information services

图 8 中空斜体数字代表 792 个互联网信息服务符合该种服务形式组合情况的数量, 其所在区域颜色越深代表落在该区域的服务数量越多。从图中可见, 在单一功能型服务中, 仅开展搜索一项形式(“1000”型)的互联网信息服务略多, 这类服务多为公开信息的查询服务, 没有仅仅提供通信或直播形式的服务; 在弱综合功能型服务中, 同时提供消息和搜索服务形式的互联网信息服务最多(y 轴正半轴处), 占

比达 79%, 其中“消息+搜索”组合形式(“1001”型)的服务占总数的 45%; 4 种服务形式都提供的强综合功能型(“1111”型)服务占总数的 14%。特别指出, 未发现 4 种服务形式都未开展的互联网信息服务(“0000”型), 说明了分类方法对互联网信息服务功能划分的完整性。

4.3.2 服务功能综合情况分析

图 9 对 792 款互联网信息服务的 4 类服务形式进行统计, 同时按强综合、弱综合和单一 3 种功能型进行累计, 可以看到弱综合功能型的服务占大多数(71%), 强综合性的服务占总数的 11%, 而单一功能型的服务仅占总数的 18%, 可见多样化功能聚合是当前互联网信息服务的主要特征。图 10 统计了 792 款互联网信息服务包含 4 类服务形式的数量分布情况, 黄色短线表示互联网站、移动应用程序和轻应用 3 种服务形态的平均数量占比, 并用柱状图形式分别展示了 3 类服务形态在 4 类服务形式的数量占比情况

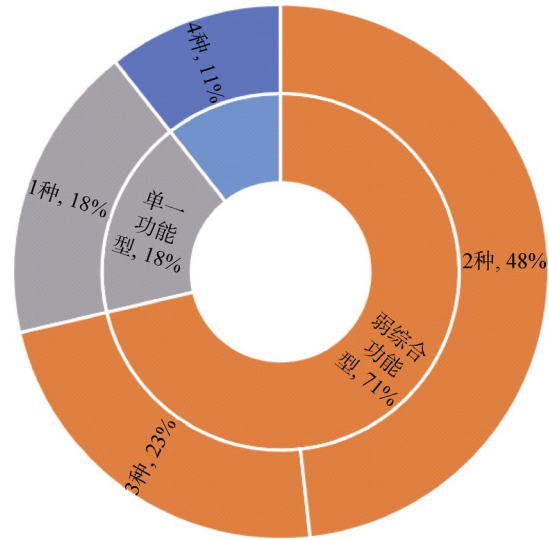


图 9 互联网信息服务的 3 种功能型分布情况
Figure 9 Distribution of three functional types of Internet information services

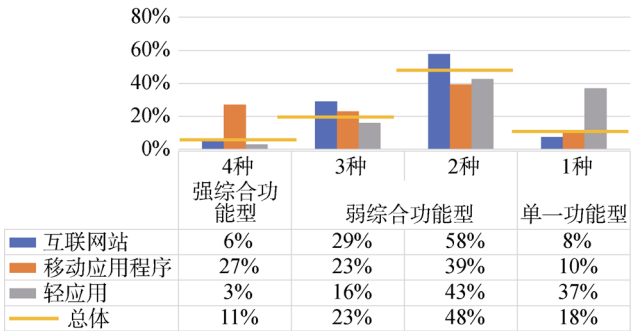


图 10 互联网信息服务的功能聚合情况
Figure 10 Internet information service function aggregation

况。从 3 类服务形态角度分析, 在包含 4 种服务形式的互联网信息服务中移动应用程序占比远高于其他 3 类, 在包含 3 种和 2 种服务形式的服务中互联网站占比最高, 而在包含 1 种服务形式的服务中轻应用占比最高, 可见移动应用程序具有鲜明的功能聚合特性, 互联网站具有的功能数量居中, 而轻应用更加侧重聚焦于单一特定功能。

5 总结

本文从互联网信息服务功能角度出发, 提出一种基于信息生命周期的自顶向下的互联网信息服务分类体系, 并从信息传播角度提出互联网信息服务分类方法, 有助于对不同服务形态、开展不同服务功能的互联网信息服务提出针对性的信息内容安全要求, 为互联网信息服务监管相关工作提供有力参考。

目前的研究工作主要集中于对互联网信息服务分类思路的探讨, 分类方法为多标签硬分类, 下一步可将分类方法推广到模糊分类方式, 从而得出互联网信息服务的主要类别建议。目前研究工作的调研范围是服务规模在行业前 1% 以上的互联网信息服务, 接下来可进一步扩大调研范围, 开展更为深入的研究。同时, 在功能分类的基础上, 我们正在研究基于服务形态、规模、功能和信息内容形式等指标的互联网信息服务信息内容安全风险分级评估方法等内容。

致谢 向本文成文中给予指导的老师、提供帮助的同学和给本文提出建议的评审专家表示衷心 □ □ □

参考文献

- [1] CNNIC. The 45th China Statistical Report on Internet Development, 2020-04-28. <http://www.cnnic.net.cn/hlwfzyj/hlwxzbg/hlwjbg/202004/P020200428596599037028.pdf>. (中国互联网络信息中心. 第 45 次中国互联网络发展状况统计报告.)
- [2] People's Daily Online. China's Mobile Internet Development (2020), 2020-07-14. <http://media.people.com.cn/BIG5/n1/2020/0714/c14677-31783128.html>. (人民网. 中国移动互联网发展报告(2020).)
- [3] We Are Social, Hootsuite. Digital 2020 Global Overview Report. 2020-01-30. <http://www.ewenku.net/docs/d5e7ecb2f3233f>.
- [4] China Internet illegal information reporting center. Acceptance of online reports in February 2020, 2020-04-30. https://www.12377.cn/tzgg/2020/bb5a2e16_web.html. (中央网信办违法和不良信息举报中心. 2020 年 4 月全国网络举报受理情况.)
- [5] Zhang B F. Research on text categorization method oriented to content security[D]. Changsha: National University of Defense Technology, 2007. (张博锋. 面向内容安全的文本分类研究[D]. 长沙: 国防科学技术大学, 2007.)
- [6] Yin J G. Determination of the Scope of Harmful Information in China's Network[J]. *Political Science and Law*, 2015(1): 102-113. (尹建国. 我国网络有害信息的范围判定[J]. *政治与法律*, 2015(1): 102-113.)
- [7] Li X Y. Study on China's Internet content regulation mechanism[D]. Wuhan: Wuhan University, 2014. (李小宇. 中国互联网内容监管机制研究[D]. 武汉: 武汉大学, 2014.)
- [8] Li J H. Discussion on the Architecture and Implementation of Hierarchical Supervision of Information Content Security[J]. *Netinfo Security*, 2001, 1(2): 22-24. (李建华. 信息内容安全分级监管的体系架构及实施探讨[J]. *信息安全*, 2001, 1(2): 22-24.)
- [9] Wen X. Establish a Hierarchical Monitoring Mechanism for Network Information Content[J]. *C-Enterprise Management*, 2010(5): 62-63. (文绪. 建立网络信息内容分级监控机制[J]. *通信企业管理*, 2010(5): 62-63.)
- [10] Yin J G. On the Rating Mechanism of Internet Information Content[J]. *Chinese Public Administration*, 2016(10): 122-127. (尹建国. 网络信息内容分级机制研究[J]. *中国行政管理*, 2016(10): 122-127.)
- [11] Zhong Y X. Principles of information science[M]. 3rd ed. Beijing: Beijing University of Posts and Telecommunications Press, 2002. (钟义信. 信息科学原理[M]. 3 版. 北京: 北京邮电大学出版社, 2002.)
- [12] Suo C J. Remarks on the Conception and Research Contents of Information Lifecycle[J]. *Library and Information Service*, 2010, 54(13): 5-9. (索传军. 试论信息生命周期的概念及研究内容[J]. *图书情报工作*, 2010, 54(13): 5-9.)
- [13] GB/T 20274. Information technology-Security techniques - Evaluation framework for information systems security assurance. 2008. (GB/T 20274. 信息技术 安全技术 信息系统安全保障评估框架. 2008)
- [14] The central people's government of the People's Republic of China. Measures for the administration of Internet information services, 2000-09-25. http://www.gov.cn/gongbao/content/2011/content_1860864.htm IDTc6TT2II0. (中华人民共和国中央人民政府. 互联网信息服务管理办法, 2000-09-25.)
- [15] Wang Q L. Study of light app development based on mobile Internet in China[D]. Guangzhou: Jinan University, 2015. (王巧莉. 我国移动互联网轻应用发展研究[D]. 广州: 暨南大学, 2015.)
- [16] Cyberspace Administration of China. Regulations on the administration of Internet news and information services, 2017-06-01. http://www.cac.gov.cn/2017-05/02/c_1120902760.htm. (国家互联网信息办公室. 互联网新闻信息服务管理规定, 2017-06-01.)

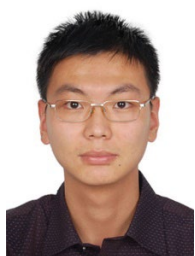
- [17] IResearch Global. PC Web Index, 2020-03 <https://index.iresearch.com.cn/pc>. (艾瑞咨询. PC Web 指数, 2020-03.)
- [18] Quest Mobile. 2019 semiannual report of small procedure, 2019-06 <https://www.questmobile.com.cn/research/report-new/56>. (Quest Mobile. 小程序 2019 半年报告, 2019-06.)
- [19] Huawei Application Market. <https://appstore.huawei.com/soft/list>. (华为应用市场.)
- [20] Xiaomi App Store. <http://app.mi.com/>. (小米应用商店.)
- [21] National Information Security Standardization Technical Committee. Notice of the national standard "information security technology Internet information service security general requirements" to solicit comments on the draft notice, 2020-01-20. https://www.tc260.org.cn/front/bzzqyjDetail.html?id=20200121161346046652&norm_id=20200112070031&recode_id=36750. (全国信息安全标准化技术委员会. 关于国家标准《信息安全技术 互联网信息服务安全通用要求》征求意见稿征求意见的通知, 2020-01-20.)
- [22] GB/T 18336. Information Technology – Security Techniques – Evaluation Criteria for IT Security[S]. 2015. (GB/T 18336. 信息技术 安全技术 信息技术安全评估准则[S]. 2015.)
- [23] Gan Q. On social mobilization[D]. Wuhan: Wuhan University, 2010. (甘泉. 社会动员论[D]. 武汉: 武汉大学, 2010.)
- [24] Safavian S R, Landgrebe D. A Survey of Decision Tree Classifier Methodology[J]. *IEEE Transactions on Systems, Man, and Cybernetics*, 1991, 21(3): 660-674.
- [25] Minto B. The Minto Pyramid Principle [M]. Democracy and Construction Press, 2006. (明托. 金字塔原理[M]. 民主与建设出版社, 2006.)
- [26] Hu Zhengrong, Duan Peng, Zhang Lei. General Theory of Communication (second edition) [M]. BEIJING: Tsinghua University Press, 2008. (胡正荣, 段鹏, 张磊. 传播学总论(第二版)[M]. 北京: 清华大学出版社, 2008.)
- [27] Schramm W, Porter W F. Men, Women, Message, and Media: Understanding Human Communication[M]. Chen Liang, et al, translate. BEIJING: Xinhua Publishing House, 1984, 123-129. (施拉姆, 波特. 传播学概论[M]. 陈亮等译. 北京: 新华出版社, 1984, 123-129.)
- [28] IResearch Global. Mobile APP Index, 2020-03. <https://index.iresearch.com.cn/new/#/app/list?type=2>. (艾瑞咨询. 移动 APP 指数, 2020-03.)
- [29] Arthur, David, and Sergi Vassilvitskii. K-means++: The Advantages of Careful Seeding[C]. *The Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, 2007, 1027-1035.
- [30] Laurens V D M, Hinton G. Visualizing Data using t-SNE[J]. *Journal of Machine Learning Research*, 2008, 9(2605): 2579-2605.
- [31] IResearch Global. Mini program Index, 2019-09. <https://index.iresearch.com.cn/new/#/miniapp>. (艾瑞咨询. 微信小程序指数, 2019-09.)



王宇航 于 2016 年在北京交通大学软件工程专业获得硕士学位, 现在中国科学院大学网络空间安全专业攻读博士学位, 现任中国科学院信息工程研究所助理研究员。研究领域为网络空间安全。研究兴趣包括信息内容安全、区块链信息服务安全、机器学习。Email: wangyuhang@iie.ac.cn



郭涛 中国科学院信息工程研究所博士生导师, 研究员, 工学博士。研究领域为网络空间安全。研究兴趣包括网络空间安全、漏洞分析与风险评估。Email: guotao@iie.ac.cn



胡静远 于 2019 年在北京大学计算机系统结构专业获得博士学位, 现在中国科学院信息工程研究所任助理研究员。研究领域包括虚拟化、内存管理、网络空间安全、舆情传播引导等。研究兴趣包括社交网络水军识别、区块链监管、智能合约漏洞分析等。电子邮件: hujingyuan@iie.ac.cn



张潇丹 于 2012 年在中国科学院大学计算机系统结构专业获得博士学位, 现任中国科学院信息工程研究所副研究员。研究领域为新型网络技术测量分析与评估。研究兴趣包括区块链、互联网数据分析、计算传播等。Email: zhangxiaodan@iie.ac.cn



韩冀中 于 2001 年在中国科学院计算技术研究所获得博士学位, 现为中国科学院信息工程研究所正高级工程师, 博士生导师。主要研究领域为大数据存储与管理、多媒体信息智能化处理。研究兴趣包括多媒体内容理解。Email: hanjizhong@iie.ac.cn