基于自然语言处理的隐私政策自动表述研究

- 技术报告 -

Janus 研究中心
上海🄵众信息技术有限公司

朱璋颖
<anonymous@pwnzen.com>

陆亦恬
<anonymous@pwnzen.com>

唐祝寿
<anonymous@pwnzen.com>

本项目受国家互联网应急中心（CERT）支持。

目录

序言

隐私政策的自动化表述是隐私政策自动化检测的基础，表述结果可用于虚假隐私政策检测、隐私政策完整性检测等方面。本文针对中文语言的特点，采用众包任务的方式对隐私政策进行标注，创建了目前为止第一个中文隐私条款训练集。使用自然语言处理技术实现了隐私政策的自动化表述工具，工具的分类模型准确率达到 *90%*。使用该工具，我们对来自华为应用市场的 *1,500* 份中文隐私政策进行了检测，检测结果表明 *38.5%* 的隐私政策为虚假隐私政策，剩余合法的隐私政策中，*92.5%* 的隐私政策在完整性方面不符合“自评估指南”的要求。在隐私政策自动表述的基础上，设计了一种隐私政策打分方法，实验结果表明大部分隐私政策的得分位于低分数区间内。

本报告是 *Janus* 研究中心在 *App* 隐私检测工作的一部分内容，报告标题、内容都将随着研究的不断深入，内容的不断扩展而变化。

Chapter 1

简介

移动应用快速发展的同时，带来了一些安全问题。移动设备做为隐私集中地，需要确保其承载的隐私信息不被移动应用滥用。

为保护用户隐私，欧盟出台了《General Data Protection Regulation》[14]（以下简称 GDPR），落实了数据控制者（App 运营者）处理数据主体（用户）信息的规则和数据主体应当享有的权利等规定。GDPR“第 29 条工作组” 还特别强调，数据控制者的应用程序应该以分层的隐私声明或通知的方式向数据主体提供隐私信息（即隐私政策）相关的链接，而不是在设备上以单一通知的形式展示此类信息。

国内也对隐私问题制定了一系列的技术规范和标准。包括：《App 违法违规收集使用个人信息自评估指南》（以下简称 “自评估指南”）、GB/T 35273《信息安全技术 个人信息安全规范》（以下简称 “安全规范”）和《信息安全技术 移动互联网应用程序（App）收集个人信息基本规范》（以下简称 “基本规范”），从隐私政策文本、收集使用个人信息行为、用户权利保障等角度对隐私政策进行了规范。

欧盟的 GDPR、国内的技术规范和标准都对隐私政策都提出了相关要求。隐私政策的目的是为了向用户说明个人信息如何被收集、使用和共享等数据实践，同时也对厂商起到约束作用，隐私政策通常可以通过链接访问的方式查看。

根据 McDonald 等人 [7] 的估计，如果认真阅读每一份隐私政策，那么身在美国的用户每年需要为此花费 201 个小时。我们的统计也表明，中文隐私政策平均包含 138 句话，用户也需要为阅读一份隐私政策花费大量的时间。隐私政策过长的篇幅、专业的内容等现实原因导致许多用户不愿意去阅读或无法直观的理解隐私政策的内容，在对内容不了解的情况下，大多直接选择接受应用的隐私政策，在这种情况下，用户对于个人信息的处理并不知情。如应用 ZAO 在其隐私政策中声明的：“在您上传及/或发布用户内容以前，您同意或者确保实际权利人同意授予 ZAO 及其关联公

司以及 ZAO 用户全球范围内完全免费、不可撤销、永久、可转授权和可再许可的权利。” 被大多数用户忽略。

针对这种现状，现有的法规/标准都对隐私政策提出了清晰易懂的要求，也有相关工作试图标准化隐私政策 [10, 3, 5, 16]。另外还有一些隐私政策自动化表述的研究工作来解决用户阅读隐私政策困难的问题。如针对英文，Polisis 等工具使用众包任务对数据进行标注、使用自然语言处理技术自动从隐私政策中提取数据实践内容 [15, 4]；CLAUDETTE [2] 使用了机器学习方法来自动检测不公平条款。

本文研究中文隐私政策的自动表述，用于定位一份隐私政策中的相关内容，在此基础上，检测虚假隐私政策，检测隐私政策的完整性。基于自动化表述的结果，我们设计了一种评分方法为隐私政策打分。

Chapter 2

工作流程

为了实现隐私政策的自动化表述，使用众包任务方式对数据进行标注，使用自然语言处理技术识别隐私政策中的相关条款。在模型建立阶段，采用众包任务方式，根据逐步优化的分类标准标注隐私政策以建立训练数据集，然后使用数据集训练分类模型，在对朴素贝叶斯、支持向量机、卷积神经网络三种分类方法比较的基础上，最终使用支持向量机对数据集进行分类；在线检测阶段中，通过分类模型对隐私政策内容进行分类，根据分类结果对隐私政策内容进行分析。具体的工作流程如图2.1所示。

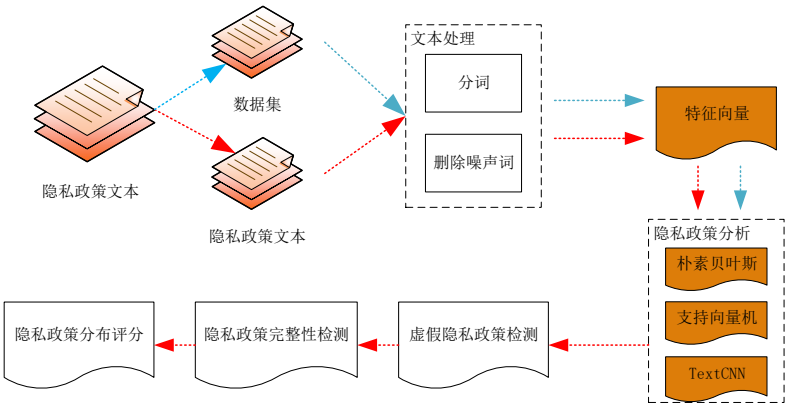
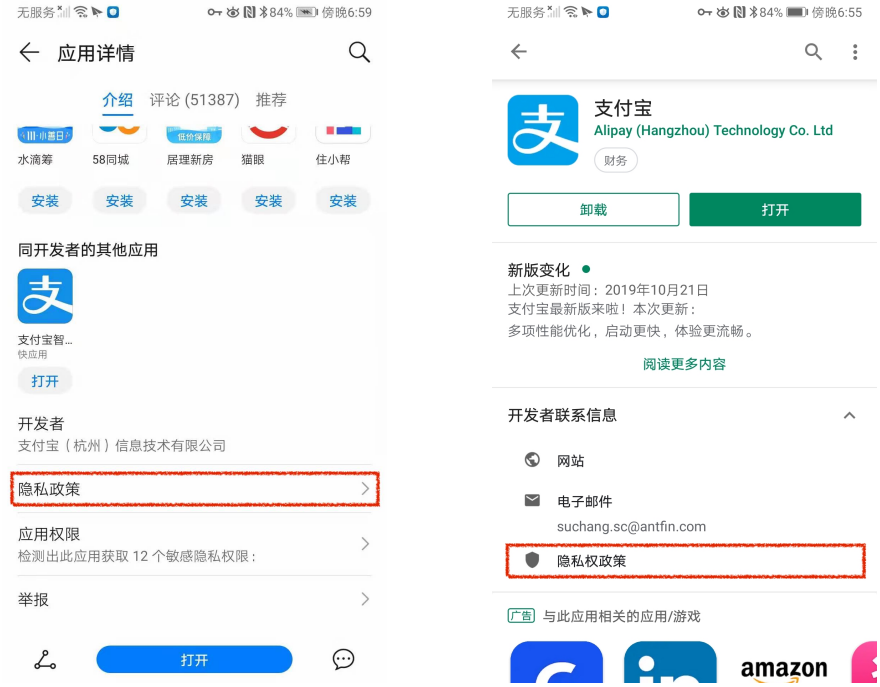


Figure 2.1: 隐私政策自动表述流程

2.1 数据集

隐私政策可以通过多种渠道采集，如搜索引擎、应用市场等。应用市场为开发者分发应用时，为开发者提供设置隐私政策链接的接口。用户在通过应用市场浏览应用时，可以通过该链接查看开发者设置的隐私政策，如图2.2所示。相比其他渠道的隐私政策，应用市场的隐私政策与移动应用紧密相关，因此质量较高。因为这些隐私政策属于公开信息，所以我们设计了针对移动应用市场的爬虫来获取这些隐私政策。具体来讲，本研究中的数据集为来源于华为应用市场的隐私政策。



(a) 华为应用市场的隐私政策　　(b) 谷歌应用的市场隐私政策

Figure 2.2: 应用市场中的隐私政策

为形成训练数据集，我们对从华为应用市场中提取的覆盖 17 种应用类型（包括影音娱乐、实用工具、社交通讯等）的 100 个热门应用的隐私政策进行了标注。接下来在 2019 年 11 月 23 日至 2019 年 11 月 28 日期间，通过持续对华为应用市场进行监控，我们爬取了 1500 份隐私政策用于检测。基于 100 篇隐私政策统计发现，平均每篇隐私政策包含 138 句话。隐私协议中句子数量的分布如图2.3所示，其中 5% 的隐私政策长度小于 50 句话，9% 的隐私政策长度大于 200 句话，隐私政策长度呈现一定的差异性。

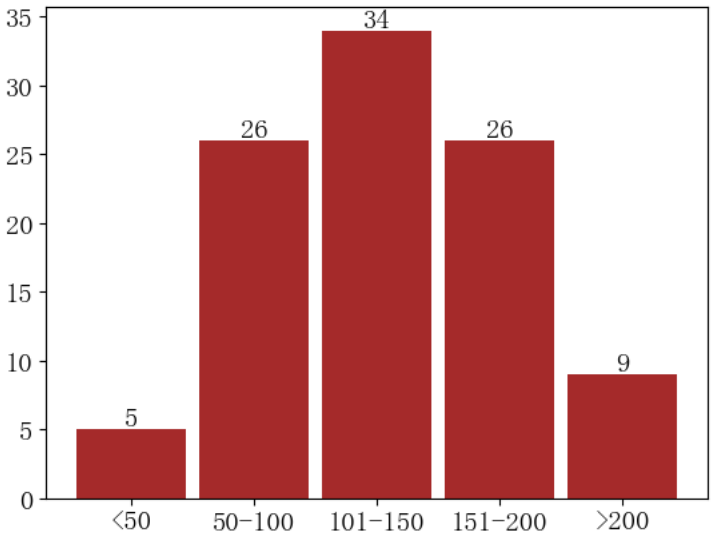


Figure 2.3: 隐私政策长度分布

2.2 数据标注

为形成训练用数据集，使用逐步优化的方法确定了标注标准、采用众包任务的方式对数据进行了标注，最终建立了带标签的数据集。该数据集是迄今为止第一个中文隐私条款训练数据集。

标注标准确定：当前法规/规范/标准比较多，包括 GDPR、“自评估指南”等。这些文件从不同的角度对隐私政策提出要求，如“自评估指南”从宏观的角度要求 APP 运营者在隐私政策文本中清晰说明个人信息规则和用户权益保障，与此同时，又从微观上对细节提出了要求，如要求 APP 运营者提供基本信息、个人信息安全保护措施和能力等。

为尽量覆盖这些文件的要求，需要建立一个可扩展的标注标准，我们借鉴 Polisis [15, 4] 的方法对隐私政策内容进行了划分，结合隐私政策进行标注过程的反馈反复进行修正，最终形成“类别-属性-值”层次结构的标注标准。该标注标准包含 7 个类别，50 个属性，91 个值[1]，部分分类标准如图2.4所示。

分类标准中的类别代表数据控制者的数据实践内容，如：第一方收集/使用、与第三方共享/转让/公开等，分别用 First-Party-Collect-Use、Third-Party-Share 等标

[1] http://ppranking.cn/static/data/China-OPP-100-Crowdsourcing-Project.pdf

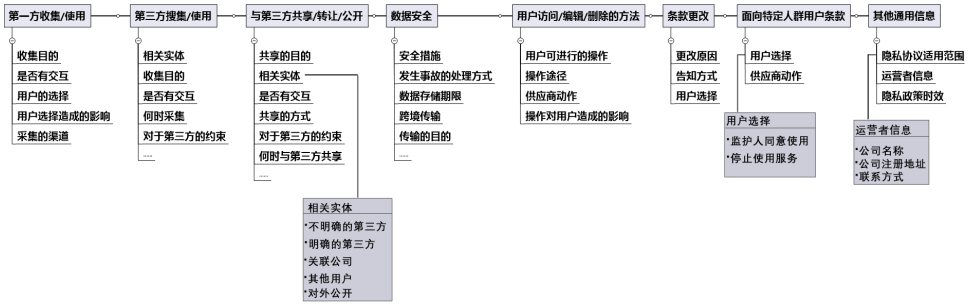


Figure 2.4: 分类标准

签表示。其他非数据实践内容，如隐私条款发布时间、生效时间被划分到其他通用信息（General-Information）类别中；属性则表示数据实践的具体内容，如将第一方收集/使用继续划分为收集目的、是否有交互、用户的选择等属性；对大部分属性，设计了相应的值选项，如为是否有交互属性设置了是和否两个值。

标注工具：标注过程使用了 BRAT [12] 这个工具。BRAT 工具可以用于解决实体识别和关系抽取等自然语言处理任务，利用该工具对文本内容添加注释，可以获得自然语言处理任务需要的标注语料。BRAT 中的标注类型分为四类，分别为实体（entities）、关系（relation）、事件（events）和属性（attributes），其中的属性绑定了一组值（values）供标注任务选择。我们利用了 BRAT 工具中的实体、属性和值字段，分别对应标注标准中的类别、属性和值字段，在此基础上，形成了如图2.5所示的在线标注工具。

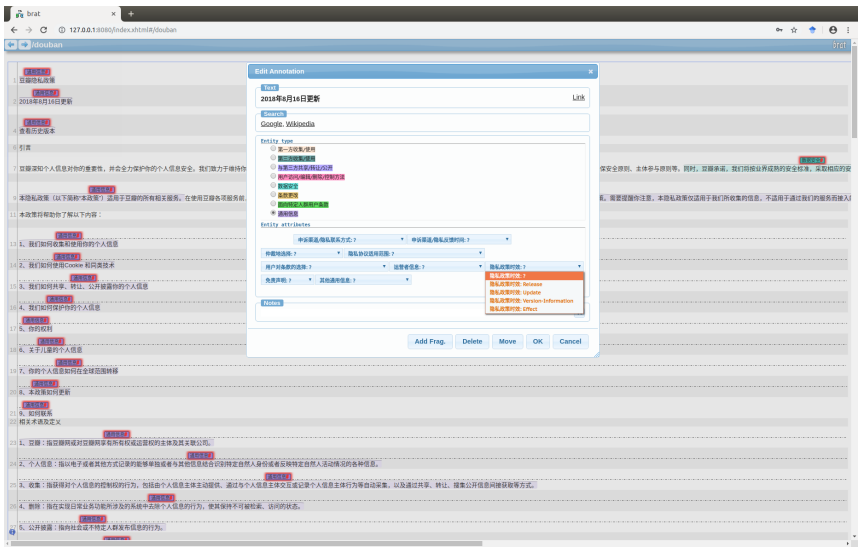


Figure 2.5: 在线标注工具

标注过程：标注过程以众包任务的方式进行，招募的 10 名学生全部来自于法律相关

专业，在确保标注者充分理解分类标准的基础上，对其开放在线标注工具入口以对隐私政策进行标注。对标注有疑问的内容都经过了充分的讨论，最终通过调整标注标准或者放弃标注等方式解决。对于每一个标签我们支付了 0.4 元的报酬，整个标注过程历时 90 天。我们通过检验数据标注的一致性，证明数据标注内容是可信的。该过程最终形成了包含 100 篇隐私政策的数据集[2]。参考了 OPP-115[15] 的命名方法，我们将该数据集命名为 Chinese-OPP-100，该数据集中共包含 11,440 个类别和属性标签。**数据处理：** 通过 BRAT 标注的结果以 ann 文件格式进行存储，如图2.6所示。ann 文件中包括所标注文本内容在隐私政策文件中的范围，所标注的类别、属性和值，每一个标签的内部编号等信息。在标注过程中，由于标签可能标注在关键词上，而分类器以句子为单位接收数据，因此将 ann 文件的内容以句子为单位进行标签合并，即如果标注内容在隐私政策中某一句话的范围内，则将其对应的类别、属性和值赋于这一句话。

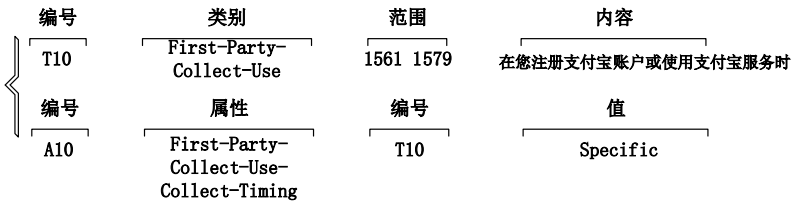


Figure 2.6: ann 格式文件内容

标注数据集按类别统计结果如表2.1所示。表2.1为 7 个类别中每个类别的标签数量，以及它们在每篇隐私政策中的均值和中位数，直观上的观察发现类别在每篇隐私政策中分布不均，First-Party-Collect-Use 在隐私政策中占比较高，说明第一方收集/使用个人信息是隐私政策中相对重要的内容。

2.3 模型训练

为能自动化表述隐私政策的类别，采用机器学习、深度学习技术对隐私政策进行分类，具体包括特征提取和模型构建工作。特征提取的目的是提取隐私政策内容的特征，将特征转化为模型可识别的格式。采用朴素贝叶斯、支持向量机、卷积神经网络三种技术构建多标签分类模型，并对分类模型进行评估。针对数据存在不均衡问题，使用惩罚学习算法。

[2]http://ppranking.cn/static/data/China-OPP-100-Crowdsourcing-Project-dataset.zip

Table 2.1: 标注数据集类别统计

类别（用标签表示）	出现次数	均值	中位数
Data-Security	598	6	7
First-Party-Collect-Use	2008	20	24
General-Information	2124	21	24
Policy-Change	212	2	2
Specific-Audience	166	2	2
Third-Party-Share	901	9	10
User-Access-Edit-Delete-Control	868	9	9

训练数据预处理：对训练数据，首先使用 JIEBA [13] 分词工具对中文文本进行分词，并删除数字、特殊符号和标点符号，同时删除诸如 “我们”，“是” 等噪声词对数据进行清洗。在此基础上，使用 TF-IDF 算法 [9] 对特征进行选择。

分类模型构建：考虑到数据实践是由多个类别组成，因此构建多标签分类模型。先对数据集中的 7 个类别构建一个多标签分类模型，之后针对每一个属性继续构建多标签分类模型，将多标签分类问题转换成多个二分类问题。具体使用了基于 SCIKIT-LEARN 工具包 [8] 的朴素贝叶斯和支持向量机，以及基于 KERAS [1] 的卷积神经网络。

(i) 朴素贝叶斯：朴素贝叶斯算法主要针对二元分类，因此采取问题转换的方法来解决隐私政策中的多标签分类问题。我们使用二元关联（BinaryRelevance，BR）分解策略 [6]，忽略标签之间的相关性，将多标签分类问题转换为多个一对多分类问题。同时构建多项式朴素贝叶斯模型来实现隐私政策文本的多标签分类。

(ii) 支持向量机：在样本数量少且特征数量多的情况下，考虑线性支持向量机，使用核函数将有限维空间映射到高维空间，使其线性可分。具体采用 SCIKIT-LEARN 工具包中 OneVsRestClassifier 进行实现，并将 kernel 参数设置为 “linear”。

(iv) 卷积神经网络：在嵌入层使用腾讯 AI Lab 的中文词向量数据 [11] 将输入内容转换为向量矩阵形式，卷积层中使用 ReLu 激活函数提取特征，经过池化层进行降维，然后在展开层，丢弃层和全连接层中进行整合处理，并防止出现过拟合现象。考虑到多标签分类，在全连接层中采用 sigmod 作为激活函数。

表2.1展示了数据不均衡问题，即一些类别的样本数量远大于其他类别的样本数量。这些不平衡数据可能导致分类模型更倾向于将新样本预测为样本数量多的类别。为了缓解不平衡数据的影响，在支持向量机和卷积神经网络中采用惩罚学习算法来

处理这个问题，使用 class_weight 参数，平衡类别之间的权重。

Table 2.2: 分类器对类别进行分类的评价指标（Precision/Recall/F1）

类别	朴素贝叶斯			支持向量机			卷积神经网络		
	精准率	召回率	F1-score	精准率	召回率	F1-score	精准率	召回率	F1-score
Data-Security	0.93	0.62	0.74	0.85	0.76	0.80	0.85	0.65	0.73
First-Party-Collect-Use	0.88	0.77	0.82	0.85	0.89	0.87	0.74	0.80	0.77
General-Information	0.91	0.61	0.73	0.87	0.85	0.86	0.89	0.80	0.77
Policy-Change	1.00	0.39	0.56	0.74	0.85	0.79	0.91	0.48	0.62
Specific-Audience	0.88	0.65	0.75	1.00	0.84	0.91	0.82	0.45	0.58
Third-Party-Share	0.91	0.62	0.74	0.71	0.81	0.76	0.78	0.78	0.78
User-Access-Edit-Delete-Control	0.91	0.56	0.69	0.86	0.83	0.84	0.83	0.70	0.76
avg	0.90	0.65	0.75	0.84	0.85	0.84	0.83	0.67	0.72

评价指标: 将数据集以 8:2 分成训练集和测试集，对于朴素贝叶斯、支持向量机和卷积神经网络，我们使用网格搜索自动调整参数。表2.2给出通过该方法在测试集上得出的类别分类器的评价指标，包括精确率、召回率和 F1 值。比较发现，朴素贝叶斯和卷积神经网络在 F1-score 指标方面没有达到与支持向量机相同的性能，支持向量机在自动化表述隐私政策的过程中表现良好。总体来讲，我们的分类器的评价指标与针对英文的自动化表述工作的 Polisis（88.4%）基本一致。附录中表A.1-A.14列出了分类器在分类每个属性时的评价指标。

Chapter 3

在线检测工具

3.1 隐私政策自动表述工具

我们对性能表现优异的支持向量机分类器进行了封装，形成了在线分析工具 [1]，如图3.1所示。该在线分析工具通过隐私政策链接爬取网页内容，对隐私政策内容进行预处理和分类，对于隐私政策自动表述结果进行着色展示，不同颜色表示不同的数据实践。

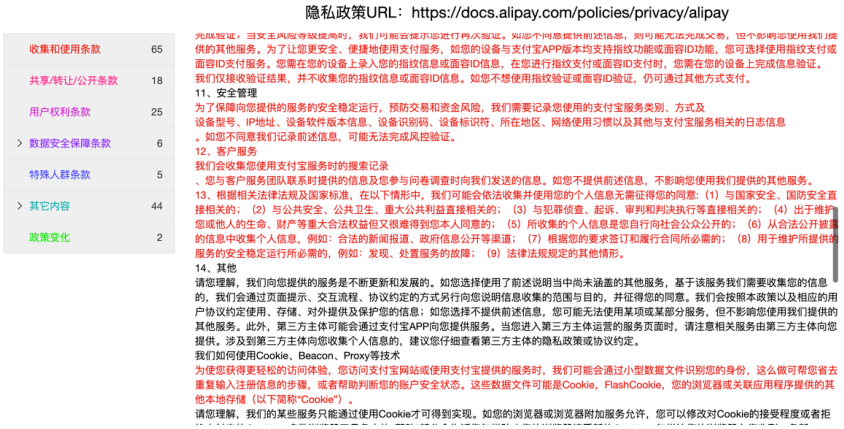


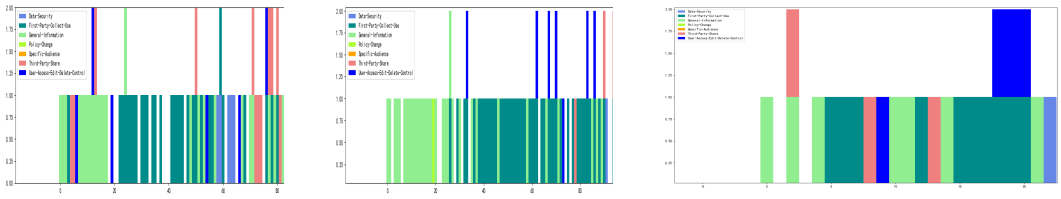
Figure 3.1: 在线分类工具

[1] http://ppranking.cn/policycheck

3.2 隐私政策图形化表示

除了在线表述工具外，我们还使用色带对隐私政策的分类结果进行图形化展示。如图3.2的三张柱状图显示了三个隐私政策分类结果的图形化表示。其中横坐标代表了隐私政策总共有多少句话，柱状图的颜色表示不同的数据实践。

实践中，我们观察到数据控制者会在隐私政策的同一句话中对多个数据实践进行描述。如在支付宝隐私政策提到：“为了保障您的信息安全，我们在收集您的信息后，将采取各种合理必要的措施保护您的信息。例如，在技术开发环境当中，我们仅使用经过去标识化处理的信息进行统计分析；对外提供研究报告时，我们将对报告中所包含的信息进行去标识化处理。我们会将去标识化后的信息与可用于恢复识别个人的信息分开存储，确保在针对去标识化信息的后续处理中不重新识别个人。”这一数据实践同时提到了第一方收集/使用和数据安全。因此，如果某一句话同时描述了多个数据实践，则在图中表现为交叠的色块。



(a) 支付宝类别分布 (b) 哔哩哔哩类别分布 (c) 作业精灵类别分布

Figure 3.2: 应用类别分布

Chapter 4

实验结果

4.1 虚假隐私政策检测

如图2.2所示，应用市场会为开发者提供填写隐私政策链接的接口，但部分开发者会在此接口处填写用户协议、官方网站或用户协议和隐私政策的混合文本链接。据我们的爬虫显示，仅华为应用市场中就存在上万个隐私链接，因此人工审核这些虚假隐私政策是不现实的。为了能自动化检测出这些虚假隐私政策，基于隐私协议自动化表述开发了检测工具。

虚假隐私政策检测是基于以下事实设计的，即隐私政策是用来描述数据实践的，即除“其他通用信息”和无法分类的信息以外的信息，如果一篇隐私政策中的数据实践内容的百分比小于 R，则认为该隐私政策为虚假隐私政策。

经过在小型数据集上的简单实验，我们将 R 设置为 0.55，并对华为市场中的 1,500 份隐私政策进行了虚假检测。在 1,500 篇被检测的文档中发现，578（38.5%）篇文档属于虚假隐私政策。随后对检测结果进行的抽样确认显示，R 的值设置是合理的，也从侧面证明数据实践内容应该占隐私政策文本的 55% 以上。

4.2 隐私政策完整性检测

如2.2节所述，目前法规/规范/标准对数据实践的完整性定义是不一致的，因此需要单独为某个法规/规范/标准来整理有针对性的检测项。本节利用数据集可扩展的优势，依据“自评估指南”的要求，来检测隐私政策的完整性。

在“自评估指南”中，要求隐私政策清晰说明个人信息处理规则及用户权益保障，其要求隐私政策所描述的内容分别对应我们设定的分类标准中的某些类别和属

性。其中和我们设定的分类标准中的类别相对应的要求包括：个人信息的使用规则、个人信息安全保护和措施能力、对外共享/转让/公开披露个人信息规则和用户权利保障机制；与属性相对应的要求包括：App 运营者信息、个人信息存储和超期处理方式、个人信息出境情况、用户申诉渠道和反馈机制、隐私政策时效和隐私政策更新。

例如，图3.2(c)所示的“应用作业精灵”在隐私政策中没有提到隐私政策更新、App 运营者信息、个人信息存储和超期处理方式、个人信息出境情况，在这些方面未能满足“自评估指南”的要求。

基于隐私政策自动化表述结果，我们对经过过滤的 922 篇合法隐私政策进行统计，如图4.1所示，从统计结果发现，总共有 853（92.5%）篇隐私政策没有完整说明“自评估指南”所要求的内容，其中有 707 篇隐私政策没有提到个人信息超期处理方式，分别有 414 篇和 447 篇没有提及个人信息存储方式和个人信息出境情况。最终实验结果表明，目前大部分隐私政策内容在完整性方面不能满足“自评估指南”的要求。

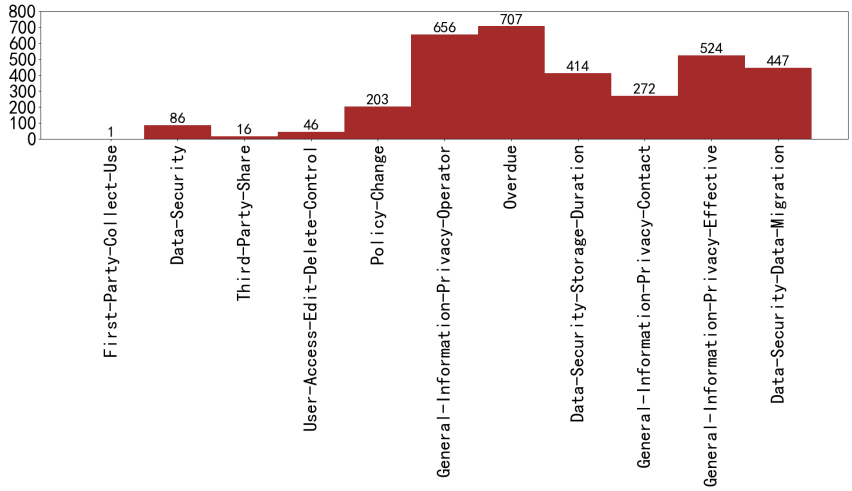


Figure 4.1: 隐私政策内容完整性检测

4.3 隐私政策的内容分布和完整性评分

观察发现，隐私政策内容分布中最常见的问题是：一个类别的内容在隐私政策中多处提及，而不同类别的描述又可能在隐私政策中某个位置交叉提及，在这种情况下，用户很难抓住数据控制者想要表达的具体实践。如很常见的“第一方收集/使用信息”和“与第三方共享/转让/公开信息”这两个类别的内容在隐私政策中交错提及，

用户容易将与第三方共享的信息误看作第一方收集使用的信息。

因此，我们认为隐私政策的内容分布越集中，则得到的评价应该越高。我们以类别为依据，对隐私政策分布情况进行评价，设计了如下评价方法：假设一篇隐私政策包含 N 句话和 M 个类别 C_i，其中某个类别 C_i 包含 n_i 句表述。针对某个类别 C_i，我们使用公式 (4.1) 的评价方法来表示其内容表述是否集中。

$$E(C_i) = \frac{N^2 - \sum_{i=1}^{n_i - 1} d_{i,i+1}^2}{N^2} \tag{4.1}$$

其中 $d_{i,i+1}$ 表示相同类别 C_i 表述内容之间的间隔，代表相同类别表述内容之间包含其他类别表述内容的数量。公式 (4.1) 的例外情况为：如果一份隐私政策中缺少与 C_i 对应的隐私政策描述，则 $E(C_i) = 0$；如果与 C_i 对应的隐私政策只有一句话，则 $E(C_i) = 1$。

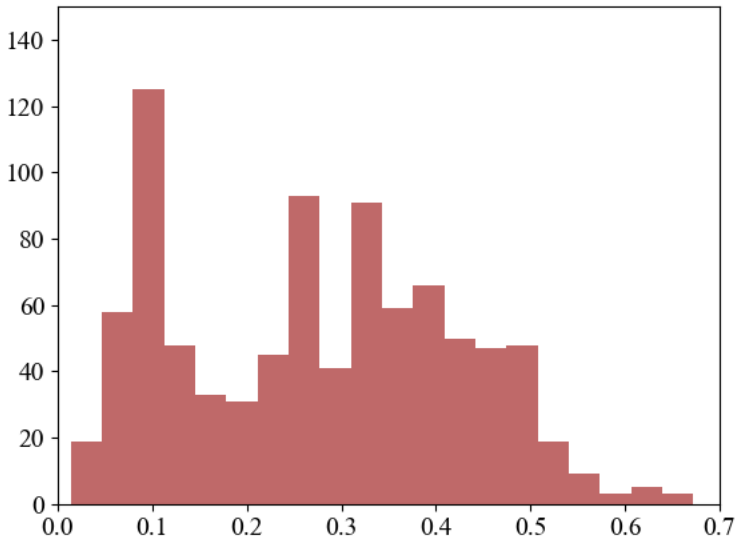
对于整篇隐私政策，使用如公式 (4.2) 评分方法。公式 (4.2) 中的 Q=138，为隐私协议中平均语句数量。之所以在类别平均评分结果上乘以系数 $\frac{N}{Q}$，是为了避免如图2.3所示的只包含少量内容的隐私政策得到高的评分，最后使用 arctan 函数进行归一化处理。

$$E = \frac{\arctan\left(\frac{\sum_{i=1}^{M} E(C_i)}{M} \times \frac{N}{Q}\right) * 2}{\pi} \tag{4.2}$$

在该评价方法下，图3.2(a)所代表的支付宝最终得分为 0.37；图3.2(b)所代表哔哩哔哩的最终得分为 0.44；作业精灵最终得分为 0.12。

基于以上算法，我们对隐私政策的内容分布作了评分，来观察隐私政策表述是否集中、内容是否完整。图4.2给出了 922 篇隐私政策的得分分布，图中横轴代表隐私政策的分数，纵轴代表评分区间内的隐私政策的数量。从中可以看出，有 32.6%(301 篇) 隐私政策在 0-0.2 区间内，6.3%(57 篇) 的隐私政策在 0.5-0.7 区间内。总体上说明集中程度较一般的隐私政策数量非常少，多数隐私政策是较好集中或是较差集中的。

Figure 4.2: 隐私政策得分分布



Chapter 5

总结

本文设计了一种隐私政策自动化表述的方法，基于该方法，我们对隐私政策的虚假性、完整性进行了评价，并对隐私政策打分。分析结果表明，目前隐私政策的总体质量较低，无法真实体现数据实践。

参考文献

[1] Francois Chollet. *Deep Learning mit Python und Keras: Das Praxis-Handbuch vom Entwickler der Keras-Bibliothek*. MITP-Verlags GmbH & Co. KG, 2018.

[2] Giuseppe Contissa et al. "Claudette meets gdpr: Automating the evaluation of privacy policies using artificial intelligence". In: (2018).

[3] Lorrie Cranor. *Web privacy with P3P*. " O'Reilly Media, Inc.", 2002.

[4] Hamza Harkous et al. "Polisis: Automated analysis and presentation of privacy policies using deep learning". In: *27th {USENIX} Security Symposium ({USENIX} Security 18)*. 2018, pp. 531–548.

[5] Patrick Gage Kelley et al. "A nutrition label for privacy". In: *Proceedings of the 5th Symposium on Usable Privacy and Security*. ACM. 2009, p. 4.

[6] Oscar Luaces et al. "Binary relevance efficacy for multilabel classification". In: *Progress in Artificial Intelligence* 1.4 (2012), pp. 303–313.

[7] Aleecia M McDonald and Lorrie Faith Cranor. "The cost of reading privacy policies". In: *Isjlp* 4 (2008), p. 543.

[8] F. Pedregosa et al. "Scikit-learn: Machine Learning in Python". In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830.

[9] Gerard Salton and Christopher Buckley. "Term-weighting approaches in automatic text retrieval". In: *Information Processing and Management: an International Journal* 24 (1988), pp. 513–523.

[10] Florian Schaub et al. "A design space for effective privacy notices". In: *Eleventh Symposium On Usable Privacy and Security ({SOUPS} 2015)*. 2015, pp. 1–17.

[11] Yan Song et al. “Directional Skip-Gram: Explicitly Distinguishing Left and Right Context for Word Embeddings”. In: *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*. New Orleans, Louisiana: Association for Computational Linguistics, June 2018, pp. 175–180. DOI: 10.18653/v1/N18-2028. URL: https://www.aclweb.org/anthology/N18-2028.

[12] Pontus Stenetorp et al. “BRAT: a web-based tool for NLP-assisted text annotation”. In: *Proceedings of the Demonstrations at the 13th Conference of the European Chapter of the Association for Computational Linguistics*. Association for Computational Linguistics. 2012, pp. 102–107.

[13] J Sun. ‘*Jieba*’ *Chinese word segmentation tool*. 2012.

[14] Paul Voigt and Axel Von dem Bussche. “The eu general data protection regulation (gdpr)”. In: *A Practical Guide, 1st Ed., Cham: Springer International Publishing* (2017).

[15] Shomir Wilson et al. “The Creation and Analysis of a Website Privacy Policy Corpus”. In: Aug. 2016, pp. 1330–1340. DOI: 10.18653/v1/P16-1126.

[16] Sebastian Zimmeck and Steven M Bellovin. “Privee: An architecture for automatically analyzing web privacy policies”. In: *23rd {USENIX} Security Symposium ({USENIX} Security 14)*. 2014, pp. 1–16.

Appendix A

附录

Table A.1: SVM for First-Party-Collect-Use

label	precision	recall	f1-score	support
Collect-Channel	0.93	0.97	0.95	108
Collect-Purpose	0.81	0.93	0.87	351
Collect-User-Choice	0.89	0.81	0.85	173
Collect-User-Choice-Influence	0.95	0.89	0.92	201
Technology-Used	0.95	0.97	0.96	168
avg	0.91	0.91	0.91	1001

Table A.2: NB for First-Party-Collect-Use

label	precision	recall	f1-score	support
Collect-Channel	0.91	0.91	0.91	108
Collect-Purpose	0.85	0.87	0.86	351
Collect-User-Choice	0.80	0.79	0.79	173
Collect-User-Choice-Influence	0.84	0.89	0.86	201
Technology-Used	0.92	0.96	0.94	168
avg	0.86	0.88	0.87	1001

Table A.3: NB for User-Access-Edit-Delete-Control

label	precision	recall	f1-score	support
Channel	0.80	0.89	0.84	265
Influence	0.96	0.89	0.93	237
Provider-Action	0.93	0.83	0.88	208
User-Action	0.79	0.80	0.80	253
avg	0.87	0.86	0.86	963

Table A.4: SVM for User-Access-Edit-Delete-Control

label	precision	recall	f1-score	support
Channel	0.80	0.87	0.84	182
Influence	0.97	0.90	0.93	168
Provider-Action	0.91	0.93	0.92	175
User-Action	0.79	0.89	0.84	180
avg	0.87	0.90	0.88	705

Table A.5: SVM for Third-Party-Collect-Use

label	precision	recall	f1-score	support
Collect-Entity	0.95	0.82	0.88	73
Collect-Interactive	0.80	0.95	0.87	43
Collect-Purpose	0.89	0.81	0.85	31
Collect-Timing	0.89	0.93	0.91	45
Enforcement	1.00	1.00	1.00	37
Technology-Used	1.00	0.90	0.95	41
avg	0.92	0.90	0.91	270

Table A.6: NB for Third-Party-Collect-Use

label	precision	recall	f1-score	support
Collect-Entity	0.90	0.81	0.85	53
Collect-Interactive	0.83	0.83	0.83	30
Collect-Purpose	0.88	0.88	0.88	25
Collect-Timing	0.79	0.97	0.87	31
Collect-Enforcement	1.00	0.96	0.98	23
Collect-Technology-Used	1.00	0.97	0.98	29
avg	0.90	0.90	0.90	191

Table A.7: NB for Data-Security

label	precision	recall	f1-score	support
Data-Migration	0.94	0.98	0.96	65
Data-Overdue	0.85	0.92	0.89	51
Event	1.00	0.92	0.96	72
Measure	0.89	0.90	0.90	105
Storage-Duration	0.92	0.80	0.85	69
avg	0.92	0.90	0.91	362

Table A.8: SVM for Data-Security

label	precision	recall	f1-score	support
Data-Migration	0.98	0.91	0.94	44
Data-Overdue	0.97	0.95	0.96	40
Event	0.98	0.94	0.96	49
Measure	0.85	0.97	0.91	89
Storage-Duration	0.92	0.77	0.84	47
avg	0.94	0.91	0.92	269

Table A.9: SVM for Policy-Change

label	precision	recall	f1-score	support
Notification	0.84	0.94	0.89	17
Reason	0.95	0.87	0.91	23
User-Choice	0.86	0.86	0.86	14
avg	0.88	0.89	0.89	54

Table A.10: NB for Policy-Change

label	precision	recall	f1-score	support
Notification	0.85	1.00	0.92	17
Reason	1.00	0.83	0.90	23
User-Choice	0.93	0.93	0.93	14
avg	0.93	0.92	0.92	54

Table A.11: SVM for Specific-Audience

label	precision	recall	f1-score	support
Response	0.97	0.87	0.92	38
User-Choice	0.94	0.98	0.96	46
avg	0.96	0.93	0.94	84

Table A.12: NB For Specific-Audience

label	precision	recall	f1-score	support
Response	0.95	0.93	0.94	42
User-Choice	0.88	0.97	0.93	39
avg	0.92	0.95	0.94	81

Table A.13: SVM For General-Information

label	precision	recall	f1-score
Operator-Information	0.95	0.97	0.96
Policy-Cover	0.93	0.85	0.96
Policy-Effectiveness	1.00	0.90	0.95
Privacy-Contact	0.92	0.59	0.72
Privacy-Response	0.97	0.86	0.91
avg	0.95	0.85	0.90

Table A.14: NB For General-Information

label	precision	recall	f1-score
Operator-Information	0.94	0.95	0.94
Policy-Cover	0.90	0.75	0.82
Policy-Effectiveness	1.00	0.75	0.82
Privacy-Contact	0.92	0.59	0.72
Privacy-Response	0.94	0.86	0.90
avg	0.93	0.80	0.86