基于 AHP-模糊综合评价法的泄露数据价值评估模型

宋 栋,张 雷,苏马婧
(华北计算机系统工程研究所,北京 100083)

摘 要:以个人隐私信息为例构建了一种泄露数据价值评估的模型。首先建立了包含不同类型数据的泄露数据集,通过抽取个人隐私相关的信息元素进行建模,利用层次分析法对各准则层的元素进行计算,得到个人隐私信息元素泄露价值的权重比例,并进行了一致性检验,通过模糊综合评价的方法对个人隐私信息泄露数据价值进行评价,得到泄露价值的评估结果。

关键词:层次分析法;个人隐私;泄露数据
中图分类号:TP399 文献标识码:A DOI:10.19358/j.issn.2096-5133.2020.09.009
引用格式:宋栋,张雷,苏马婧. 基于 AHP-模糊综合评价法的泄露数据价值评估模型[J].信息技术与网络安全,2020,39(9):44-48.

Value evaluation model of leaked data based on AHP-fuzzy comprehensive evaluation method

Song Dong, Zhang Lei, Su Majing
(National Computer System Engineering Research Institute of China, Beijing 100083, China)

Abstract: This article uses personal privacy information as an example to construct a model for evaluating the value of leaked data. Firstly, a leaked data set containing different types of data is established. By extracting personal privacy-related information elements for modeling and calculation, the elements of each criterion layer are calculated using analytic hierarchy process to obtain the weighted proportion of personal privacy information element leakage value. The consistency test is carried out, and the value of the leakage of personal privacy information is evaluated by the method of fuzzy comprehensive evaluation, and the evaluation result of the leakage value is obtained.

Key words: analytic hierarchy process(AHP); personal privacy; leaked data

0 引言

随着社会的高速发展,发达的科技使得信息流通更加便利,人们之间的交流越来越频繁。由于大数据的应用越来越广泛,大量数据带来巨大价值的同时也带来了数据被泄露的风险。

在日常生活和办工、上网购物、网络社交的过程中,不可避免需要用户提供个人信息或其他相关信息,在这过程中用户的个人隐私信息就不可避免地存在被泄露的风险。数据泄露一般由数据存储设备被盗窃、网络攻击泄露、个人疏忽或失误泄露、企业内部人员行为泄露以及通过勒索软件泄露等原因导致。重要数据的泄露对企业或个人带来的影响十分严重,而目前对泄露数据带来风险和危害的评估研究却比较少。本文以个人隐私信息数据为例通过层次分析法和模糊综合评价的方法对个人隐私信息泄露数据的价值进行评估,从而对评估不同种类泄露数据的价值提供参考。

1 层次分析法的基本原理

层次分析法(Analytic Hierarchy Process, AHP)是20世纪70年代初美国运筹学家 SAATY T L 提出的,是美国国防部研究“根据各个工业部门对国家福利的贡献大小进行电力分配”课题时,应用网络系统理论和多目标综合评价方法,提出的一种层次权重决策分析方法。它是一种将与决策总是有关的元素分解成目标、准则、方案等层次,在此基础上进行定性和定量分析的决策方法。该方法将定量分

析与定性分析结合起来，用求解判断矩阵特征向量的办法，求得每一层次的各元素对上一层次某元素的优先权重，然后再用加权和的方法递阶归并各备择方案对总目标的最终权重，最终权重最大者即为最优方案。

1.1 层次分析法的一般步骤[1-2]

(1)根据情况建立层次结构模型

根据实际问题，将影响最终结果的相关因素采用自上而下的方式进行分层，分别为目标层、准则层(或指标层)、对象层(或方案层)，同一层之间的因素基本上相对独立，上层因素受下层因素的影响。形成层次分析法的分析结构模型。

(2)构造判断(成对比较)矩阵

根据 1-9 标度法和成对比较法，构造两两对比矩阵，用同等重要、稍微重要、重要、很重要以及非常重要等判断表示各层上的每一个因素两两对比的情况。

(3)层次单排序及进行一致性检验

对每一层两两比较矩阵计算最大特征根和特征向量，然后进行一致性检验，如果一致性检验结果正常，则可将特征向量作为该层的权向量。

(4)层次总排序及进行一致性检验

通过组合下层的权向量，对总层次进行排序并进行一致性检验，若通过，则说明按此排序符合要求。

1.2 层次分析法的特点

(1)系统性：可将评估的对象比作系统，按照分解、比较、判断、综合的思维方式进行分析与决策；

(2)实用性：通过对评估对象进行定性与定量相结合的方式进行评估，较传统的方法更易做出决策；同时，层次分析法的应用范围很广，可以应用于经济计划和管理、生产决策、交通运输等领域，可以处理决策、评价、分析和预测等类型的问题。

2 利用 AHP 对个人隐私信息泄露数据进行权重分析

2.1 建立泄露数据集

假设给定一个泄露数据集 Ω，其中包含 N 条泄露数据，记为 $d_1, d_2, \cdots, d_N$，每条泄露数据包含 M 个信息元素，记为 $e_1, e_2, \cdots, e_M$。在泄露数据集 Ω 中抽取个人隐私信息作为一条泄露数据 d_1，包含个人身份证号码、电话号码、家庭住址、性别、付款记录信息、收货地址、健康数据信息、医疗记录、微信账号和密码、微博账号和密码等信息元素。信息元素越重要，包含这些信息元素的个人隐私信息泄露后对个人带来的危害就越高，即泄露数据的价值就越高。

2.2 确定个人隐私信息泄露数据价值的权重

(1)建立个人隐私信息泄露数据价值体系

本文将个人隐私信息的元素分为四个类别，分别为个人基本信息、个人购物及支付信息、个人医疗信息以及个人社交网络信息，从而建立如图 1 所示的个人隐私信息泄露数据价值体系。从图中可知，目标层 A 为个人隐私信息泄露数据价值，准则层 B 代表个人隐私信息的四个类别，子准则层 C 为个人隐私信息的具体元素。通过建立个人隐私信息泄露数据价值体系可以对个人隐私信息泄露数据价值进行评估。

(2)对第二层指标构建两两比较矩阵 A-B[3]

对准则层的任意两个指标 a_i 和 a_j 进行重要性比较的赋值[4-5]，如表 1 所示。

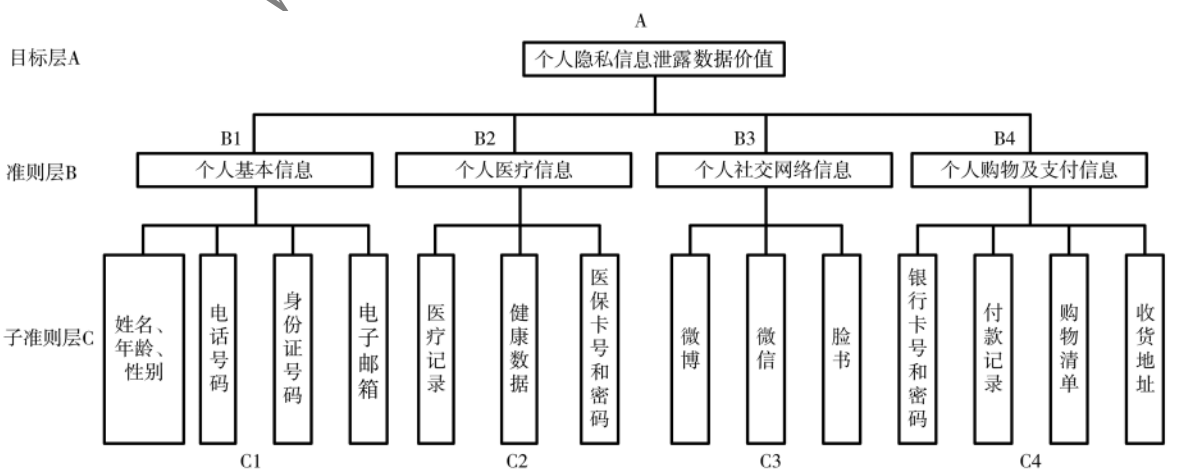

图 1 个人隐私信息泄露数据价值体系

表 1 重要性结果赋值

重要性赋值	重要性比较
1	a_i 和 a_j 一样重要
3	a_i 比 a_j 稍微重要
5	a_i 比 a_j 明显重要
7	a_i 比 a_j 非常重要
9	a_i 比 a_j 特别重要
1/3	a_i 比 a_j 稍微不重要
1/5	a_i 比 a_j 明显不重要
1/7	a_i 比 a_j 非常不重要
1/9	a_i 比 a_j 特别不重要

采用 1-9 标度法对四类信息进行两两对比，形成目标层 A 对准则层 B 的两两比较情况，两两比较矩阵如表 2 所示。

表 2 A-B 比较情况

	个人基本信息	个人医疗信息	社交网络信息	购物及支付信息
个人基本信息	1	7/5	7/3	7/5
个人医疗信息	5/7	1	5/3	1
社交网络信息	3/7	3/5	1	3/5
购物及支付信息	5/7	1	5/3	1

因此可知 A-B 的比较矩阵 A 为：

$$A=[a_{ij}]=\begin{bmatrix}1 & 1.4 & 2.33 & 1.4\\ 0.71 & 1 & 1.67 & 1\\ 0.43 & 0.6 & 1 & 0.6\\ 0.71 & 1 & 1.67 & 1\end{bmatrix} \tag{1}$$

对矩阵 A 按列进行归一化可得：

$$A1=[a1_{ij}]=\frac{a_{ij}}{\sum_{i=1}^{n}a_{ij}}$$

$$=\begin{bmatrix}0.35 & 0.35 & 0.345 & 0.35\\ 0.249 & 0.25 & 0.247 & 0.25\\ 0.15 & 0.15 & 0.148 & 0.15\\ 0.249 & 0.25 & 0.247 & 0.25\end{bmatrix} \tag{2}$$

对 $A1$ 按行求和可得：

$$V=[v_i]=\sum_{j=1}^{n}a1_{ij}=\begin{bmatrix}1.395\\ 0.996\\ 0.598\\ 0.996\end{bmatrix} \tag{3}$$

因此可求得 A-B 的权重为：

$$W_A=[w_i]=\frac{v_i}{\sum_{i=1}^{n}v_i}=\begin{bmatrix}0.348\\ 0.294\\ 0.149\,5\\ 0.294\end{bmatrix} \tag{4}$$

计算比较矩阵 A 的赋权和向量：

$$AW=\begin{bmatrix}1 & 1.4 & 2.33 & 1.4\\ 0.71 & 1 & 1.67 & 1\\ 0.43 & 0.6 & 1 & 0.6\\ 0.71 & 1 & 1.67 & 1\end{bmatrix}\begin{bmatrix}0.348\\ 0.294\\ 0.149\,5\\ 0.294\end{bmatrix}$$

$$=\begin{bmatrix}1.519\\ 1.085\\ 0.652\\ 1.085\end{bmatrix} \tag{5}$$

所以计算比较矩阵 A 的最大特征根：

$$\lambda_{max}=\sum_{i}\frac{(AW)_i}{nW_i}=4.026\,7 \tag{6}$$

式中，n 为矩阵的阶数，$i,j=1,2,\cdots,n$。

(3)进行一致性检验

计算一致性指标 CI[6]：

$$CI=\frac{\lambda_{max}-n}{n-1}=0.008\,92 \tag{7}$$

计算一致性率 CR：

$$CR=\frac{CI}{RI}=\frac{0.008\,92}{0.89}0.010\,02 \tag{8}$$

其中，RI 为随机一致性指标[7]，通过查表可知当矩阵的阶数为 4 时，对应的 RI=0.89。

通过一致性率 CR=0.010 02<0.1 可知，对比矩阵 A 的一致性是可以接受的。

通过同样的方法对子标准层各元素的权重进行计算，分别为：

$$W_{c1}=\begin{bmatrix}0.056\\ 0.167\\ 0.5\\ 0.28\end{bmatrix},\lambda_{c1max}=4.079$$

$CR_{c1}=0.027<0.1$，因此符合一致性检验。

$$W_{c2}=\begin{bmatrix}0.152\,8\\ 0.354\,1\\ 0.492\,9\end{bmatrix},\lambda_{c2max}=3.005\,1$$

$CR_{c2}=0.004\,4<0.1$，因此符合一致性检验。

$$W_{c3}=\begin{bmatrix}0.238\,8\\ 0.623\,7\\ 0.137\,3\end{bmatrix},\lambda_{c3max}=3.015\,3$$

$CR_{c3}=0.026\,4<0.1$，因此符合一致性检验。

$$\boldsymbol{W}_{c4}=\begin{bmatrix}0.547\\0.116\ 3\\0.060\ 2\\0.280\ 4\end{bmatrix},\lambda_{c4\max}=4.047$$

$CR_{c4}=0.017\ 5<0.1$，因此符合一致性检验。

由上述结果可知，个人隐私信息泄露数据价值体系权重如表 3 所示。

表 3 个人隐私信息泄露数据价值体系权重表

目标层 A	准则层 B		子准则层 C	
	准则	权重	子准则	权重
个人隐私信息泄露数据价值	个人基本信息	0.348	姓名、年龄、性别	0.056
			电话号码	0.167
			身份证号码	0.5
			电子邮箱	0.28
	个人医疗信息	0.294	医疗记录	0.152 8
			健康数据	0.354 1
			医保卡号和密码	0.492 9
	个人社交网络信息	0.149 5	微博	0.238 8
			微信	0.623 7
			脸书	0.137 3
	个人购物及支付信息	0.294	银行卡号和密码	0.547
			付款记录	0.116 3
			购物清单	0.060 2
			收货地址	0.280 4

2.3 使用模糊综合评价法对个人隐私信息泄露数据价值进行综合评估

利用层次分析法对个人隐私信息各元素进行了权重的分析，最后通过模糊综合评价方法对个人隐私信息泄露数据的价值进行评估。

(1)建立模糊评价矩阵

假设将个人隐私信息泄露数据的价值分为四个等级，即 $V=\{$非常高，高，一般，低$\}$。对 20 个人进行个人隐私信息泄露数据价值评价的问答，根据每个人对个人隐私信息包含的元素信息的不同评价可得到模糊评价矩阵，如针对个人基本信息包含的信息元素的模糊评价矩阵，归一化后为：

$$\boldsymbol{R}_{c1}=\begin{bmatrix}0.05&0.1&0.25&0.6\\0.3&0.4&0.3&0\\0.9&0.1&0&0\\0.5&0.25&0.25&0\end{bmatrix}\tag{9}$$

(2)计算个人隐私信息泄露数据评价矩阵

计算二级模糊综合评价矩阵 $\boldsymbol{B}_{c1}$ 为：

$$\begin{aligned}\boldsymbol{B}_{c1}&=\boldsymbol{w}_{c1}\cdot\boldsymbol{R}_{c1}\\&=[0.056\quad 0.167\quad 0.5\quad 0.28]\begin{bmatrix}0.05&0.1&0.25&0.6\\0.3&0.4&0.3&0\\0.9&0.1&0&0\\0.5&0.25&0.25&0\end{bmatrix}\\&=[0.642\ 9\quad 0.189\quad 0.134\quad 0.034]\end{aligned}\tag{10}$$

同理可知：

$$\begin{aligned}\boldsymbol{B}_{c2}&=\boldsymbol{w}_{c2}\cdot\boldsymbol{R}_{c2}\\&=[0.152\ 8\quad 0.354\ 1\quad 0.492\ 9]\begin{bmatrix}0.25&0.6&0.15&0\\0.75&0.25&0&0\\0.9&0.1&0&0\end{bmatrix}\\&=[0.747\quad 0.229\quad 0.023\quad 0]\end{aligned}\tag{11}$$

$$\begin{aligned}\boldsymbol{B}_{c3}&=\boldsymbol{w}_{c3}\cdot\boldsymbol{R}_{c3}\\&=[0.238\ 8\quad 0.623\ 7\quad 0.137\ 3]\begin{bmatrix}0.3&0.5&0.2&0\\0.5&0.4&0.1&0\\0.15&0.35&0.4&0\end{bmatrix}\\&=[0.404\quad 0.417\quad 0.165\quad 0]\end{aligned}\tag{12}$$

$$\begin{aligned}\boldsymbol{B}_{c4}&=\boldsymbol{w}_{c4}\cdot\boldsymbol{R}_{c4}\\&=[0.547\ \ 0.116\ 3\ \ 0.060\ \ 0.280]\begin{bmatrix}1&0&0&0\\0.25&0.25&0.5&0\\0&0&0.4&0.6\\0.4&0.3&0.3&0\end{bmatrix}\\&=[0.688\quad 0.113\quad 0.166\quad 0.036]\end{aligned}\tag{13}$$

因此，一级模糊评价矩阵为：

$$\boldsymbol{R}_A=\begin{bmatrix}0.642\ 9&0.189&0.134&0.034\\0.747&0.229&0.023&0\\0.404&0.417&0.165&0\\0.688&0.113&0.166&0.036\end{bmatrix}\tag{14}$$

由此可知个人隐私信息泄露数据评价矩阵 $\boldsymbol{B}$ 为：

$$\boldsymbol{B}=\boldsymbol{w}_A\cdot\boldsymbol{R}_A=[0.706\quad 0.229\quad 0.127\quad 0.022]\tag{15}$$

(3)评估结果

通过计算可知，对于以个人基本信息、个人医疗信息、社交网络信息以及购物及支付信息为信息元素的个人隐私泄露数据，70.6%的人认为价值为非常高，22.9%的人认为价值为高，12.7%的人认为价值为一般，只有 2.2%的人认为价值为低。根据最大隶属度原则[8-9]可知，对此条个人隐私信息泄露数据价值的评价为非常高。

利用层次分析法和模糊综合评价的方法可以对包含不同信息元素的泄露数据的重要性进行评估，从而可对不同种类的泄露数据进行重要性排序

并对泄露数据价值进行评估。

3 防止泄露数据事件发生及个人隐私保护的建议

随着网络技术的飞速发展，隐私安全成为了每个人生活中不可忽略的重要部分，怎样能够有效地保护个人的隐私信息不被泄露成为了当前非常重要的一个话题。本文针对防止泄露数据事件发生及个人隐私保护提出了一些建议，如下：

(1)为了减少泄露数据事件的发生，应当采取相应的技术防护措施，如对数据或文件进行加密处理，定期进行电脑漏洞检测并及时进行修复等。

(2)对于企业，加强企业员工内部管理是减少企业泄露数据事件发生的重要途经，如对员工的上网行为进行审计、监测和管控并对数据文件传输进行权限管理。

(3)对于个人隐私信息，首先，应在自我意识上提高重视程度，不要轻易将个人隐私信息告诉陌生人；其次，在浏览网页的时候，不要轻易将个人隐私信息暴露在网络环境中；再者，在使用智能手机时，不要轻易下载未知的 APP 进行使用；最后，在使用个人电脑时应当避免设备自动连接到公共网络。

4 结论

本文主要研究泄露数据价值的评估，以个人隐私信息为例，通过使用层次分析法和模糊综合评价的方法对个人隐私信息泄露数据进行了权重分析和价值评估，得出以个人基本信息、个人医疗信息、个人社交网络信息以及个人购物及支付信息为信息元素的个人隐私信息泄露数据的价值为非常高的结论。同理通过分析和计算可以对包含不同信息元素的泄露数据的重要性进行评价，进而对不同种类的泄露数据的价值进行评估。

参考文献

[1] 钱颂迪.运筹学[M].北京：清华大学出版社，2000.

[2] 韩利，梅强，陆玉梅，等.AHP-模糊综合评价方法的分析与研究[J].中国安全科学学报，2004，14(7)：86-89.

[3] 胡庆宇.基于层次分析法的大学生综合素质多级模糊评价[D].保定：华北电力大学，2010.

[4] 张吉军.模糊层次分析法(FAHP)[J].模糊系统与数学，2000，14(2)：80-88.

[5] 杨雯雯，李庆春.层次分析法与模糊综合评价法在高校学生评教问题中的应用[J].白城师范学院学报，2017(10)：34-37.

[6] 牛泽力.大学生综合素质评价模型的研究[D].保定：华北电力大学，2013.

[7] 彭建国.基于模糊综合评判的大学生综合评价与实证研究[D].成都：成都理工大学，2009.

[8] 吴秉坚.模糊数学及其经济分析[M].北京：中国标准出版社，1994.

[9] 张再罗，张伟.模糊综合评价法在高校教师授课质量评价中的应用[J].科技与创新管理，2014，35(1)：58-61.

(收稿日期：2020-05-10)

作者简介：

宋栋(1986-)，男，硕士，工程师，主要研究方向：网络测量、工控及网络安全。

张雷(1985-)，男，硕士，工程师，主要研究方向：社交网络、自然语言处理、工控安全。

苏马婧(1985-)，女，博士，高级工程师，主要研究方向：网络探测。

版权声明