基于特征分析的智能网联汽车数据分级方法研究

冀智华，王　瑞，张　巧
（中汽智联技术有限公司，天津　300300）

摘　要：当前，智能网联技术的发展使得汽车成为数据交互的重要载体，智能网联汽车所产生的数据量呈现指数型增长趋势，国家对于数据安全的重视程度也在不断加强。在此背景下，从行业实践角度出发，对我国当前数据安全监管现状、数据分级方法进行梳理，最后提出一种基于特征分析的数据分级方法，助力企业加强数据治理、满足政府合规要求。

关键词：智能网联汽车；数据分级；特征分析；层级分析
中图分类号：TP311.13　　**文献标识码：**A　　**DOI：**10.19358/j.issn.2097-1788.2024.04.011
引用格式：冀智华，王瑞，张巧. 基于特征分析的智能网联汽车数据分级方法研究[J]. 网络安全与数据治理，2024，43(4)：67-70.

Research on data classification and grading methods for intelligent connected vehicles based on feature analysis

Ji Zhihua，Wang Rui，Zhang Qiao
（CATARC Intelligent and Connected Technology Co.，Ltd.，Tianjin 300300，China）

Abstract：Currently，the development of intelligent connected technology has made automobiles become an important carrier of data exchange. The amount of data generated by intelligent connected vehicles is showing an exponential growth trend，and the country's emphasis on data security is also constantly strengthening. In this context，this article，from the perspective of industry practice，sorts out the current situation of data security supervision and data classification and grading methods in China. Finally，a data grading method based on feature analysis is proposed to assist enterprises in strengthening data governance and meeting government compliance requirements.

Key words：intelligent connected vehicle；data classification；feature analysis；analytical hierarchy process

0　引言

智能网联技术的发展推动着汽车由传统交通工具转变为集成影音娱乐、智能导航、驾驶辅助等功能的智能移动载体，驾乘人员在使用智能网联汽车过程中产生大量交互数据[1]。未来，一辆L4级自动驾驶汽车每8 h产生数据预计将达到10 TB，高级别自动驾驶将进一步推动单车数据指数级增加。同时，L2级及以上乘用车渗透率逐年提升，2022年全年渗透率达到29.4%，2023年一季度进一步提升至33.4%，年新增数量达千万辆级[2]。无论从汽车数量还是单车数据产生量，都将迎来快速增长，海量数据应运而生。

国家对于数据的重视程度也在不断提升。2022年12月19日，中共中央、国务院印发《关于构建数据基础制度更好发挥数据要素作用的意见》，提出加快构建数据基础制度，激活数据要素潜能；2023年2月27日，中共中央、国务院印发的《数字中国建设整体布局规划》中强调释放数据要素价值，赋能经济社会发展，数据价值与重要性日益提升[3]。同时，《中华人民共和国数据安全法》中明确规定国家建立数据分类分级保护制度，对数据要实施分类分级保护；包括智能网联汽车产品准入、车联网安全检查等工作均对数据安全提出相关要求。因此，对汽车数据开展分类分级不仅有助于汽车企业挖掘海量数据商业价值，同时可帮助企业进行数据安全风险管控，满足政府合规要求[4]。

1　研究现状

1.1　政策法规层面

法律层面，《中华人民共和国网络安全法》指出网络运营者应当采取数据分类、重要数据备份和加密等措施，

以保障网络数据不被泄露或者被窃取、篡改；《中华人民共和国数据安全法》提出根据数据在经济社会发展中的重要程度，以及一旦遭到篡改、破坏、泄露或者非法获取、非法利用，对国家安全、公共利益或者个人、组织合法权益造成的危害程度，对数据实行分类分级保护。《中华人民共和国个人信息保护法》提出个人信息处理者要对个人信息实行分类管理，以防止未经授权的访问以及个人信息泄露、篡改、丢失。三部法律均对数据分类或分级提出相关要求，各有侧重又相互交叉，为后续汽车行业规章、制度、标准的出台奠定基础[5-6]。

部门规章层面，我国汽车行业主管部门对汽车生产、销售、使用等不同环节进行管理。《工业和信息化领域数据安全管理办法（试行）》对工业和信息化领域数据分类分级、重要数据和核心数据识别认定、备案等提出管理要求，并提出将工业和信息化领域数据分为一般数据、重要数据和核心数据三级。《汽车数据安全管理若干规定（试行）》中定义了汽车数据中的个人信息、敏感个人信息、重要数据以及汽车数据处理者的含义和类型，并明确了汽车在设计、生产、销售、使用、运维等过程中的涉及个人信息数据和重要数据的处理原则，为数据分类分级与重要数据识别提出客观要求[7]。

指导意见层面，工业和信息化部装备工业一司发布的《智能网联汽车生产企业及产品准入管理指南（试行）（征求意见稿）》以及工业和信息化部等四部门发布的《关于开展智能网联汽车准入和上路通行试点工作的通知》文件中均提出应建立智能网联汽车产品数据资产管理台账，实施数据分类分级管理，加强个人信息与重要数据保护。国家互联网信息办公室发布的《数据出境安全评估申报指南（第一版）》指出数据处理者应当建立包括分类分级、个人信息权益保护等在内的数据安全保障能力。从顶层法律到部门规章，再到行业指导意见，我国对数据分类分级均提出相关要求[8-9]。

1.2 标准层面

全国汽车标准化技术委员会制定的《智能网联汽车数据通用要求》从危害性（国家安全、公共利益、个人权益）和重要性（数据处理者达成预设目标的关键程度、数据处理者潜在利益、数据处理者投入成本）两方面对数据分级进行评估确定，将数据划分为 S0～S3 共 4 级。

金融、电信与互联网行业因其数据业务较为集中，数据类型相对简单，目前在分类分级方面完成度相对较高。其中，金融行业《金融数据安全 数据安全分级指南》提出，安全性（保密性、完整性、可用性）为数据分类分级的重要参考属性，同时考虑影响对象与影响程度要素将数据分为 5 级[10]。电信行业《基础电信企业数据分类分级方法》根据基础电信企业移动支付业务数据重要程度与敏感程度以及泄露后对国家安全、社会秩序、企业经营管理和公众利益造成的影响和危害程度，并结合保密性、完整性、可用性三个属性遭破坏后造成的影响将数据分为 4 级[11]。全国信息安全标准化技术委员会发布的《信息安全技术 网络数据分类分级规则》则从整体影响对象、影响程度两个维度对一般、重要、核心数据进行定义，并给出具体识别规则。

1.3 实践层面

现有智能网联汽车数据分级方法主要根据数据危害程度与重要性进行定级，主要包括：正则表达式、语种类型匹配、数据内容分析、机器学习模型训练等。其中正则表达式方法主要应用于在不同数据的组合中识别重要或敏感数据字段，以判定数据的重要程度，其针对数据本身属性变化进行分级能力相对不足。语种类型匹配与数据内容分析两种方法主要通过语言逻辑与概率图模型等相结合的方式进行数据分级匹配，对同一类型数据不同场景难以有效识别。机器学习模型通过决策树、向量机等算法训练，对离散的数据类型进行分析，但该类算法不能充分考虑数据的动态性及数据间的关联特性，难以形成准确分析。以上几类方法针对数据“时效性、关联性与精度”等特性未进行动态识别，同时，存在对智能网联汽车数据定义不一致问题。

因此，本文提出一种基于特征分析的智能网联汽车数据分级方法（Intelligent Connected Vehicle Data Classification Method Based on Feature Analysis，ICV-DCFA）。该方法将数据“时效性、关联性、精度”特征作为定级要素进行识别；基于定级要素，结合影响对象、影响范围、影响程度参考维度，形成数据定级基本规则[12]。

2 ICV-DCFA 计算过程

基于特征分析的数据分级方法通过利用层次分析法，将影响对象、影响范围、影响程度，以及数据时效性、关联性、精度三个特征作为判断数据级别的六个基本准则，通过定性分析转化为定量决策的方式，使数据分级的结果更加准确。层次分析确定数据级别的流程如图 1 所示。

2.1 构建数据分级层次结构模型

数据分级层次结构模型主要包含目标层、准则层与方案层三层。其中，目标层为要解决的问题；准则层为要达到目标需要关注和对比的准则；方案层为具体解决方案。在 ICV-DCFA 中，将所划分的数据级别（如 1 级，2 级，3 级……）作为可选方案层，结合影响对象、影响范围、影响程度，以及数据时效性、关联性、精度特征进行计算，最后得出数据重要程度的最优解。具体数据分级层次结构模型如图 2 所示。

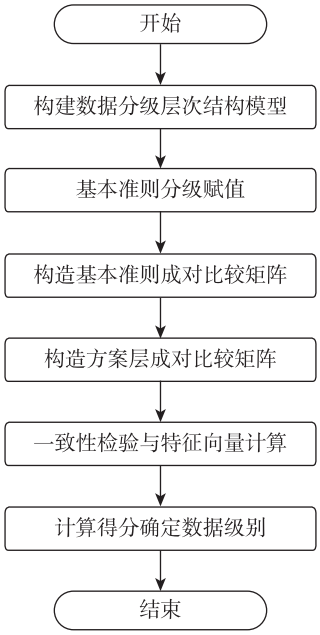

图 1　数据级别确定流程图

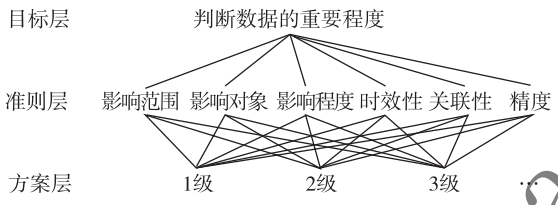

图 2　数据分级层次结构模型

2.2　基本准则分级赋值

如前所述，数据分级的六个基本准则包括：影响对象、影响范围、影响程度三个影响要素，以及时效性、关联性、精度三个数据特征。其中，影响对象包括国家安全、公共利益/社会秩序、用户/个人；影响范围包括行业、企业、机构；影响程度包括非常严重、严重、中等、轻微、无。

数据的时效性、关联性及精度需要按照国家法律法规相关要求及企业实际业务中数据间的关联关系确定。如《中华人民共和国网络安全法》第二十一条规定，“采取监测、记录网络运行状态、网络安全事件的技术措施，并按照规定留存相关的网络日志不少于六个月。”则与网安法要求相对应的数据在六个月时间内，其重要程度相对较高，超出法律规定最短时间则重要性降低。

设基本准则为 p_i（$1 \leqslant i \leqslant 6$），则每个准则子项为 p_{ik}（$1 \leqslant k \leqslant n$），$n$ 为每个准则的子项数。通过对数据基本准则进行梳理，并按照数据实际情况对准则内各子项进行赋值。为与层次分析法赋值范围保持一致，p_{ik} 的赋值范围为［1，9］。

2.3　构造基本准则成对比较矩阵

设基于基本准则生成的比较矩阵为 $\boldsymbol{C}$，如表 1 所示。

表 1　基本准则成对比较矩阵

	p_1	p_2	p_3	p_4	p_5	p_6
p_1	1	p_1/p_2	…	…	…	…
p_2	p_2/p_1	1	…	…	…	…
p_3	…	…	1	…	…	…
p_4	…	…	…	1	…	…
p_5	…	…	…	…	1	
p_6	…	…	…	…	…	1

设比较矩阵元素为 C_{ij}，则：

$$C_{ij} = \begin{cases} 1, & i = j \\ p_i/p_j, & i \neq j \end{cases} \tag{1}$$

2.4　构造方案层成对比较矩阵

设方案层元素，即数据级别为 l_i（$1 \leqslant i \leqslant n$），$n$ 为数据最高级。则方案层成对比较矩阵构建如表 2 所示。

表 2　方案层成对比较矩阵

p_1	l_1	l_2	…	l_n
l_1	1	l_1/l_2	…	…
l_2	l_2/l_1	1	…	…
…	…	…	1	
l_n	…	…	…	1
		…		
p_6	l_1	l_2	…	l_n
l_1	1	l_1/l_2	…	…
l_2	l_2/l_1	1	…	…
…	…	…	1	
l_n	…	…	…	1

数据每一项基本准则都分别于方案层构建成对比较矩阵，共构建 6 个比较矩阵。

2.5　一致性检验与特征向量计算

对生成的成对比较矩阵进行一致性检验，以保证各元素重要度之间的协调性，避免出现 a 比 b 重要，b 比 c 重要，c 又比 a 重要的矛盾情况出现。

成对比较矩阵 $\boldsymbol{C}$ 的一致性检验计算公式为：

$$\mathrm{CI} = \frac{\lambda_{\max}(\boldsymbol{C}) - n}{n-1} \tag{2}$$

$$\mathrm{CR} = \frac{\mathrm{CI}}{\mathrm{RI}} \tag{3}$$

其中，$\lambda_{\max}(\boldsymbol{C})$ 为矩阵 $\boldsymbol{C}$ 的最大特征根，n 为比较矩阵阶数，CR 为一致性比率，RI 为平均一致性指标。

RI 值与成对比较矩阵阶数对应关系如表 3 所示。

表 3　RI 与比较矩阵阶数对应表

1	2	3	4	5	6	7	8	9
0. 00	0. 00	0. 58	0. 90	1. 12	1. 24	1. 32	1. 41	1. 45

当 CR < 0. 1 时，判定成对比较矩阵 $\boldsymbol{C}$ 具有满意的一致性，或其不一致程度是可以接受的；否则检查成对比较矩阵 $\boldsymbol{C}$ 中各元素赋值并修正。

成对比较矩阵一致性检验通过后计算各矩阵对应的特征向量并进行归一化处理。

2. 6　确认数据级别

设基于基本准则成对比较矩阵 $\boldsymbol{C}$ 的特征向量为 $\boldsymbol{\omega}_C$，方案层数据级别 l_i 成对比较矩阵 $\boldsymbol{P}_i$ 对应特征向量为 $\boldsymbol{\omega}_i$，数据重要度向量为 $\boldsymbol{W}$。则 $\boldsymbol{W}$ 计算公式如下：

$$W_i = \boldsymbol{\omega}_i \times [\boldsymbol{\omega}_1, \boldsymbol{\omega}_2, \cdots, \boldsymbol{\omega}_n] \tag{4}$$

$$\boldsymbol{W} = [W_1, W_2, \cdots, W_n] \tag{5}$$

其中，W_1，W_2，…，W_n 表示重要程度，与数据级别 l_1，l_2，…，l_n 相对应。max［W_1，W_2，…，W_n］即为该项数据对应级别。

3　ICV-DCFA 分析

3. 1　ICV-DCFA 优点

本文提出的 ICV-DCFA 方法具有以下优点：

（1）ICV-DCFA 方法在传统影响对象、影响范围、影响程度基础上，新增时效性、关联性与精度三个定级要素，可以更加全面准确地评估数据的重要程度。同时，时效性、关联性与精度三个定级要素具有很强的动态特性，这与实际数据的动态变化相契合。

（2）ICV-DCFA 方法通过引入层级分析法 AHP，将定性分析转化为定量分析，保证评价尺度的一致性，结果更加精准。同时，该方法有助于通过自动化工具自主完成计算，提高数据分级的效率。

（3）数据级别随着数据特征变化而动态变化，利用 ICV-DCFA 方法，仅需提前设置好相应规则，数据分级结果即可自行调整，免去重新计算评估过程，有助于企业数据分级管理。

3. 2　ICV-DCFA 实际应用

智能网联汽车数据种类丰富，数据特征多，本方法中选择时效性、关联性、精度三个特征作为定级要素，实践过程中，企业也可结合实际应用需求丰富完善 ICV-DCFA 基本准则的选择，以更加全面准确确定数据重要程度。

在选定基本准则后需要根据数据实际情况进一步确认子级并赋值，同时需要根据数据动态性设置相应规则。

ICV-DCFA 方案层及数据级别未做具体明确划分。《工业和信息化领域数据安全管理办法》将数据分为一般数据、重要数据、核心数据，企业可根据实际情况划分数据登记，并运用 ICV-DCFA 计算重要程度，确认数据级别。

4　结论

本文从汽车行业数据分类分级实际需求出发，提出了基于特征分析的智能网联汽车数据分级方法 ICV-DCFA。利用层次分析法结合数据时效性、关联性、精度等特征，将有助于通过定量计算方式推动企业数据分级能力建设，并动态调整数据重要程度，以满足合规要求。未来工作，将基于企业实践参数，进一步确定数据特征及等级等参数，以形成更加准确的评价模型，进一步推动智能网联汽车数据分级落地实践。

参考文献

[1] YU Z Y, CAI K X. Perceived risks toward in-vehicle infotainment data services on intelligent connected vehicles [J]. Systems, 2022, 10 (5): 162 - 162.

[2] 孙逢春. 充分发挥大数据效能 提升新能源汽车安全水平 [J]. 智能网联汽车, 2022 (3): 10 - 12.

[3] 朱千一. 论数据分类分级的目的、原则与规则 [C]//上海市法学会. 智慧法治学术共同体文集, 2023.

[4] 王会杰, 杨燕红, 李志强. 我国智能网联汽车发展现状及策略分析 [J]. 汽车实用技术. 2023, 48 (6): 53 - 57.

[5] 徐子森. 智能网联汽车数据处理的法律规制：现实、挑战及进路 [J]. 兰州大学学报 (社会科学版), 2022, 50 (2): 100 - 111.

[6] 吴海燕, 陈朴, 陈亚亮, 等. 智能网联汽车数据安全国内外治理机制及政策研究 [J]. 电信快报, 2022 (9): 27 - 33.

[7] 周亮, 张晓娟, 邱意民, 等. 电力数据分类分级方法研究 [J]. 电力信息与通信技术, 2023, 21 (4): 25 - 30.

[8] 张巧, 李翠萍, 边旭东, 等. 基于能力成熟度模型的车联网漏洞管理探索 [J]. 中国信息安全, 2023 (4): 88 - 91.

[9] 李旸. 高校图书馆网络安全风险评估模型构建 [J]. 网络安全技术与应用, 2023 (3): 78 - 80.

[10] 全国金融标准化技术委员会. 金融数据安全 数据安全分级指南：JR/T 0197—2020 [S]. 2020.

[11] 中国通信标准化协会. 基础电信企业数据分类分级方法：YD/T 3813 - 2020 [S]. 2020.

[12] MA Y L. Research on safety risk assessment method of wind power enterprises based on hybrid analytic hierarchy process [J]. Journal of Physics: Conference Series, 2023 (1): 1 - 6.

（收稿日期：2024 - 02 - 02）

作者简介：

冀智华（1996 - ），男，硕士研究生，工程师，主要研究方向：数据安全、数据治理。

王瑞（1990 - ），男，硕士研究生，工程师，主要研究方向：数据安全、数据治理。

张巧（1992 - ），女，硕士研究生，工程师，主要研究方向：网络安全、数据安全。

版权声明

凡《网络安全与数据治理》录用的文章，如作者没有关于汇编权、翻译权、印刷权及电子版的复制权、信息网络传播权与发行权等版权的特殊声明，即视作该文章署名作者同意将该文章的汇编权、翻译权、印刷权及电子版的复制权、信息网络传播权与发行权授予本刊，本刊有权授权本刊合作数据库、合作媒体等合作伙伴使用。同时，本刊支付的稿酬已包含上述使用的费用，特此声明。

《网络安全与数据治理》编辑部