基于模糊影响图的差分隐私算法保护效果评估方法

田月池[1,2,3]，李凤华[1,2,3]，周泽峻[1,2,3]，孙哲[4]，郭守坤[1,3]，牛犇[1,3]
（1.中国科学院信息工程研究所，北京 100085；2.中国科学院大学网络空间安全学院，北京 100049；
3.网络空间安全防御重点实验室，北京 100085；4.广州大学网络空间安全学院，广东 广州 510006）

摘　要：针对隐私保护算法实际保护效果评估难的问题，提出了一种基于模糊影响图的差分隐私算法保护效果评估方法，实现对差分隐私算法的多维度评估，得出保护效果综合分数和等级。从算法安全性、算法可行性、隐私偏差性、数据可用性和用户体验5个方面出发，建立指标体系。使用模糊理论处理不确定性，通过影响图传递影响关系并计算该模糊影响图，得出保护效果分数和等级，据此反馈调整算法参数，实现迭代评估。提出正向化环节，解决截然相反的算法在某些情况下评估结果一样的问题。电–碳模型中的对比实验表明，所提方法能够对差分隐私算法的保护效果做出有效评价，消融实验进一步表明，正向化环节对算法的区分度起了关键作用。

关键词：隐私保护效果；综合评估；模糊影响图；差分隐私
中图分类号：TN92
文献标志码：A
DOI：10.11959/j.issn.1000–436x.2024122

Assessment method on protection effectiveness of differential privacy algorithms based on fuzzy influence diagram

TIAN Yuechi[1,2,3], LI Fenghua[1,2,3], ZHOU Zejun[1,2,3], SUN Zhe[4], GUO Shoukun[1,3], NIU Ben[1,3]
1. Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100085, China
2. School of Cyber Security, University of Chinese Academy of Sciences, Beijing 100049, China
3. Key Laboratory of Cyberspace Security Defense, Beijing 100085, China
4. Cyberspace Institute of Advanced Technology, Guangzhou University, Guangzhou 510006, China

Abstract: In response to the challenge of comprehensively assessing privacy-preserving algorithms, an assessment method on protection effectiveness of differential privacy algorithms based on fuzzy influence diagram was proposed, achieving a multi-perspective assessment of differential privacy algorithms with a comprehensive score and level as assessment results. Starting from five aspects—algorithm security, feasibility, privacy bias, data utility, and user experience, an indicator system was established. Fuzzy theory was employed to handle uncertainties, while the diagram was used to propagate interactions between factors. The assessment score and level were obtained by calculating the fuzzy influence diagram, and then used as feedback for parameter adjustment to achieve iterative assessment. Formalization link was proposed to solve the problem of completely opposite algorithms with idential evaluation results. Comparative experiments on electricity-carbon analysis model demonstrate the proposed method can assess the protection effectiveness of differential privacy algorithms effectively. Ablation experiments further show that the formalization link plays a key role in the discrimination of the algorithm.

Keywords: privacy protection effectiveness, comprehensive assessment, fuzzy influence diagram, differential privacy

收稿日期：2024–02–23；**修回日期：**2024–06–11
通信作者：郭守坤，guoshoukun@iie.ac.cn
基金项目：国家重点研发计划基金资助项目（No.2021YFB3100300）；国家自然科学基金资助项目（No.62332018）；国家社科基金重大项目（No.22&ZD147）
Foundation Items: The National Key Research and Development Program of China (No.2021YFB3100300), The National Natural Science Foundation of China (No.62332018), Major Programs of the National Social Science Foundation of China (No.22&ZD147）

0 引言

随着通信技术、网络技术和计算技术的持续发展和广泛应用，数字经济成为当今社会不可或缺的经济形态，数据作为现代社会的新型生产要素，在频繁跨域流通、泛在共享的过程中蕴含的大量用户隐私信息被采集、留存、交换、衍生，如何对其在全生命周期进行有效保护成为备受关注的研究热点。不断涌现的隐私保护方案若使用不当仍会导致隐私泄露，因此对保护效果的有效评估至关重要，我国制定了一系列法律法规和标准体系对评估工作做了规范。除了满足合规性要求以外，有效评估有助于发现隐私保护方案潜在的安全漏洞以提高整体安全性，还可以帮助识别现有算法的局限性，推动隐私保护技术发展。另外，对隐私保护的效果进行评估并公开结果可以提高组织的透明度，有助于提升用户对服务提供商的信任。

隐私计算[1]作为面向隐私信息全生命周期保护的计算理论和方法，是隐私信息的所有权、管理权和使用权分离时隐私度量、隐私泄露代价、隐私保护与隐私分析复杂性的可计算模型与公理化系统，从计算角度出发，考虑隐私具有与时间、空间、信息类型、主体主观特性等密切关联的特点，涵盖了信息搜集者、发布者和使用者在信息产生、发布、存储、使用、销毁等全生命周期的所有计算操作。

隐私计算框架在隐私信息全生命周期的各个环节中建立应用场景、保护需求与计算模型之间的映射关系，如图1所示。其中，隐私效果评估环节作为隐私计算的重要抓手，根据相关评价准则，判断对隐私信息的保护效果是否达到用户或法规要求，支撑算法以及算法管理方案的迭代修正，若所选择的算法达不到效果，则向之前环节进行反馈迭代，直至达到效果。

差分隐私通过向数据中添加噪声，使得攻击者无法区分数据集是否包含用户的某条隐私记录，从而保护用户隐私。然而，差分隐私在达到保护隐私目的的同时，也带来了额外的误差，对数据可用性造成损害，为了综合考量这些影响，需要对差分隐私保护算法进行综合评估。

现有的隐私保护效果评估方案大多关注在单一指标上的表现，使用基于信息论的度量[2-6]、基于概率论的度量[7-9]、基于攻击效果的指标[10-11]、基于隐私预算下界的指标[12-16]等来衡量隐私保护的效果，评价维度较为单一。使用多个指标的评估方案[17-20]大多只能给出一系列孤立结果，不能给出一个能直观反映算法整体隐私保护效果的综合结果。少数给出综合结果的评估方案[21-22]侧重比较指标的相对重要性和指标之间的相互作用，不能很好地处理各指标之间可能存在的相互影响、相互耦合的现象。

模糊影响图[23]是一种基于不确定信息解决复杂评估问题的图形模型，已在多个领域被用于解决实际问题。基于此，本文提出一种基于模糊影响图的差分隐私算法保护效果评估方法，以解决现有评估方案评价维度单一、忽略指标间相互作用、无法合成综合评估结果等问题。具体地，本文提出针对差分隐私算法的通用评估指标体系，对差分隐私算法进行定量与定性结合的多维评估，同时考虑了隐私的主观性，能够处理受主观认知或偏好影响的、不易量化的问题的评估；利用影响弧表示相互制约关系，充分显示差分隐私方案影响因素间的关系；利用影响图传递模糊推理过程，给出考虑多因素后的综合评估结果，反映算法的整体保护效果。

本文主要的贡献总结如下。

1) 构建了多维度的评估指标体系，支持多种应用场景、多种类型的差分隐私算法的效果评估，具有通用性，且考虑了指标间相互制约、交叉影响关系。

2) 提出了基于模糊影响图的差分隐私算法保护效果评估方法。使用模糊逻辑量化指标值、使用

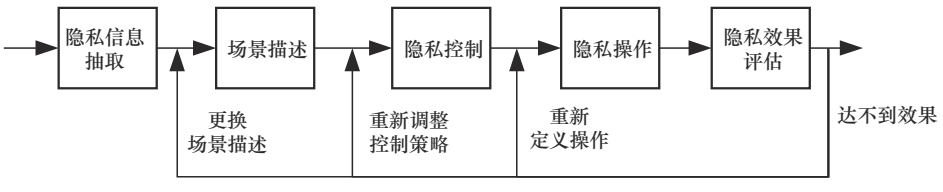

图1 隐私计算框架

影响弧描述指标间的影响关系，得到定性和定量的隐私保护效果综合评估结果。

3) 根据评估结果反馈迭代，为按需更换场景描述、重新调整控制策略、重新定义操作等提供指导，支撑隐私信息全生命周期的保护。

1　相关工作

1.1　隐私保护效果评估

现有的针对隐私保护效果开展的评估方案大多关注单一指标，如使用基于信息论的度量[2-6]、基于概率论的度量[7-9]、基于攻击效果的指标[10-11]、基于隐私预算下界的指标[12-16]等来衡量隐私保护的效果。

以信息论为基础的度量常用于评估隐私泄露情况。张文静等[2]用真实位置和单个扰动位置之间的互信息度量单一时刻位置的隐私性，用真实位置和多个扰动位置之间的互信息度量攻击者掌握多个扰动位置时的隐私泄露。Zamani 等[3]将互信息作为度量指标，使用隐私数据与发布数据之间的互信息度量隐私泄露风险，使用可用数据与发布数据之间的互信息度量可用性，从而实现二者的均衡。针对差分隐私算法的评估，Cuff 等[4]使用互信息描述攻击者可获取到的信息量，利用其与隐私预算之间的大小关系表征该隐私保护机制的正确性和隐私泄露风险。Jagielski 等[5]提出了一种基于隐私损失值的差分隐私保护效果评估方法。该方法从差分隐私的数学定义出发，将隐私损失值作为隐私强度的量化指标，评估差分隐私算法对整个训练集的保护效果。通过结合统计学的方法，评估出在一定置信度前提下的隐私损失值下界，从而实现量化隐私保护强度的目的。Wang 等[6]分析了几种不同的基于 f 散度的信息论指标在隐私保护中的作用和限制并给出了解释，然后基于 χ^2 散度开发了数据驱动的隐私风险度量框架。

从统计学角度出发，以概率论为基础的度量也可以用于评估数据经保护后的隐私性或可用性。文献[7-9]使用概率论相关指标分别衡量隐私性和可用性，并试图找到二者最优平衡点。Asoodeh 等[7-8]计算隐私性约束下根据加噪数据正确推断原始数据的概率，将其作为隐私性的度量。Rassouli 等[9]针对数据发布场景提出了基于概率质量函数的隐私泄露度量方案，并使用互信息、最小均方误差和误差概率度量可用性。

攻击效果也可以用于评估隐私保护方案的效果。为了评估松弛差分算法隐私泄露的风险，Jayaraman 等[10]将“隐私泄露率”引入差分隐私保护效果评估中，提出了一种基于成员推理攻击和属性推理攻击的差分隐私保护效果评估方法。该方法以基于梯度扰乱机制的差分隐私算法为评估对象，分析了不同松弛差分隐私算法对隐私保护效果的影响。Ryu 等[11]系统性地研究了差分隐私对协同推理的保护效果，通过对 4 种数据集上的 3 种协同推理模型使用重构攻击，使用均方误差、平均结构相似性、峰值信噪比等指标度量重构攻击的效果，进而评估差分隐私的保护效果。

为了保证差分隐私算法的正确性，对于差分隐私算法不满足预期隐私保护能力的问题，主要关注其隐私预算下界的情况。Ding 等[12]基于假设检验，启发式地搜索违反差分隐私的示例。Bichsel 等[13]提出一种基于概率论的差分隐私保护能力量化方案，基于采样和优化等方法，逼近各类随机差分扰动算法的隐私预算下界。在此基础上，Bichsel 等[14]利用二分类器缩小搜索空间，但计算出的边界并不紧密，为此 Niu 等[15]通过最优化方法改进了边界的紧密程度。此外，Askin 等[16]通过新的启发式方法缩小了搜索空间，并用概率密度函数替代了以频率估计概率的方法，提高了搜索效率和精度。

上述方案实现了对所关注指标的准确评估，但评价维度较为单一，评估结果并不全面。

文献[17-20]使用多个指标，多角度地对隐私保护方案进行评估。Sakib 等[17]使用平均主观损失评估攻击者基于公开信息猜测正确的概率，使用平均客观损失评估信息泄露程度，使用平均置信度提升评估攻击者通过观察额外数据对其猜测正确性的信心水平的整体改善。Ye 等[18]使用基于相对误差和均方误差的差分隐私保护效果评估方法，考虑差分隐私算法的准确性、数据规模适应性、数据相关性、通信开销和时间开销等指标。Oya 等[19]提出一种位置隐私保护优化机制，结合条件熵和最坏情况数据质量损失等多种指标，利用线性规划的方法寻找最佳隐私保护机制。Wang 等[20]提出一种针对差分隐私保护效果的度量评价方案 CheckDP，可以对满足差分隐私的机制自动生成

证明，并对不满足所声称隐私预算的差分隐私机制生成反例。

上述方案未给出能直观反映算法整体隐私保护效果的综合评估结果，虽然在评估过程中考虑了多个指标，但只关注算法在各个指标上的独立表现并给出孤立结果，不能指导决策者对多个不同的算法进行比较。

多维综合评估方法[21-22]通过选取适宜的若干评估指标并赋予合理权重，计算出基于多个维度的综合评估结果。Chen等[21]从App中提取6类影响隐私风险的因素，使用模糊聚类和知识依赖理论为每个因素分配合适的权重，得出定性和定量的风险评估结果。俞艺涵等[22]考虑影响网络环境下数据安全风险的10个因素，并基于模糊综合评价的方法对数据分区块、多层次地评估。

上述方案要求不同评估指标相互独立、互不相关，因此该类方案侧重比较指标的相对重要性，没有考虑因素间的交叉影响和相互依赖制约关系，不能很好地处理各指标之间可能存在的相互影响、相互耦合的现象。

1.2 基于模糊影响图的评估

模糊影响图是一种基于不确定信息解决复杂评估问题的图形模型，由刘金兰等[23]综合应用模糊集理论与影响图理论而提出，并应用在工程项目风险分析系统中，解决了以概率论为基础的传统方法无法包含大量影响项目的随机因素的问题。对于算法中独立节点频率矩阵的计算方法，程铁信等[24]做了改进和完善，并分析其数学思想。此后，模糊影响图方法在诸多领域的广泛应用中不断发展完善。

刘玉梅等[25]使用模糊影响图对高速列车传动系统可靠性的影响因素进行分类分析，通过概率分布的相近程度，识别影响传动系统可靠性的关键因素。门金柱等[26]利用模块化的思想，将舰载直升机作战效能分为各个模块，评估各种影响因素对作战效能的综合影响以及具体如何影响各个模块，分析影响作战效能的关键因素。Xia等[27]采用模糊影响图实现对战争结果的宏观预测和风险评估，通过分别计算战胜和战败概率的期望，预测战争结果。

模糊影响图理论虽然已广泛应用于多个领域，但在隐私保护效果评估领域的使用仍较少，需要修改完善以适应该领域的特点并有效应用。

2 预备知识

2.1 差分隐私

2个仅相差一条记录的数据集称为相邻数据集，当差分隐私保证分别对这2个数据集进行同一查询访问时，产生同一结果的概率接近1。而松弛型差分隐私放宽该约束条件，允许差分隐私以一定的概率失效。具体地，松弛型差分隐私定义如下。

(ε,δ)-差分隐私。设有随机算法M，P_M为M所有可能的输出构成的集合。对于任意2个相邻数据集D和D'以及P_M的任意子集S_M，若算法M满足

$$\Pr[M(D)\in S_M]\leqslant \mathrm{e}^{\varepsilon}\Pr[M(D')\in S_M]+\delta$$

则称算法M满足(ε,δ)-差分隐私（$\varepsilon>0,\delta\in[0,1]$），其中，非负参数$\varepsilon$为隐私预算，其大小直接限制上述概率，而该限制以最大δ的概率不成立。

全局敏感度。设有查询函数$f:D\rightarrow R^d$，对于任意2个相邻数据集D和D'，若

$$\Delta=\max_{D,D'}\left\|f(D)-f(D')\right\|_1$$

则称为查询函数f的全局敏感度，其中，$\left\|f(D)-f(D')\right\|_1$是$f(D)$和$f(D')$之间的1-阶范数距离。全局敏感度是决定差分隐私算法加入噪声量大小的关键参数。

2.2 影响图

影响图作为描述随机变量和决策之间依赖关系的方法，由Howard等[28]提出。影响图是由节点和有向边组成的无环有向图，其中，圆形代表机会节点，表示所研究问题中的随机变量，菱形代表价值节点，表示被建模系统的价值或所要求的产出；有向边代表影响弧，表示节点间的相互影响，从影响因素节点指向被影响因素节点。根据节点状态是否受其他节点的影响，节点又可分为无入度的独立节点和有入度的非独立节点。

2.3 模糊集合及运算

Zadeh[29]将经典数学的应用范围扩展到了模糊领域，提供了一种处理不确定性和不精确性问题的新方法。

1) 模糊集和隶属函数

论域：被讨论的对象的全体，用大写字母U表示。

模糊集：设U为论域，称从U到$[0,1]$的映射

$$\mu_A:U\to[0,1],\quad x\mapsto\mu_A(x)\in[0,1]$$

确定了一个U上的模糊子集A；称映射μ_A为A的隶属函数，$\mu_A(x)$表示元素x隶属于A的程度，即x对A的隶属度。

2) 模糊集合的并集和交集

设A和B是论域U上的2个模糊子集，则A和B的并集$A\cup B$及交集$A\cap B$的隶属函数分别为

$$\mu_{A\cup B}(x)=\mu_A(x)\vee\mu_B(x)=\max\left(\mu_A(x),\mu_B(x)\right)$$

$$\mu_{A\cap B}(x)=\mu_A(x)\wedge\mu_B(x)=\min\left(\mu_A(x),\mu_B(x)\right)$$

3) 模糊关系及合成

模糊关系：设论域U和V，则称集合$U\times V=\{(u,v)|u\in U,v\in V\}$为笛卡儿积。$U\times V$上的一个模糊子集$\tilde{R}$为从集合$U$到集合$V$的模糊关系，由它的隶属函数

$$\mu\tilde{R}:U\times V\to[0,1],\quad(u,v)\mapsto\mu\tilde{R}(u,v)\in[0,1]$$

确定，记作

$$U\xrightarrow{\tilde{R}}V$$

其中，隶属度$\mu\tilde{R}(u,v)$表示u与v具有关系$\tilde{R}$的程度，其计算方法为$\mu\tilde{R}(u,v)=\mu_U(u)\mu_V(v)$。

对于有限论域，模糊关系可用矩阵表示。设A和B分别是有限论域U和V上的2个模糊子集，设A包含n个元素$\{a_1,a_2,\cdots,a_n|a_i\in U,1\leqslant i\leqslant n\}$，$B$包含$m$个元素$\{b_1,b_2,\cdots,b_m|b_m\in V,1\leqslant i\leqslant m\}$，将这些元素的隶属度分别组成2个向量$\left[\mu_A(a_1),\mu_A(a_2),\cdots,\mu_A(a_n)\right]$和$\left[\mu_B(b_1),\mu_B(b_2),\cdots,\mu_B(b_m)\right]$，则$U$到$V$的一个模糊关系$A\times B$为上述2个向量的外积所得到的矩阵

$$A\times B=\left[\mu_A(a_1),\mu_A(a_2),\cdots,\mu_A(a_n)\right]\otimes\left[\mu_B(b_1),\mu_B(b_2),\cdots,\mu_B(b_m)\right]=\begin{bmatrix}\mu_A(a_1)\mu_B(b_1) & \cdots & \mu_A(a_1)\mu_B(b_m)\\ \vdots & \ddots & \vdots\\ \mu_A(a_n)\mu_B(b_1) & \cdots & \mu_A(a_n)\mu_B(b_m)\end{bmatrix}$$

模糊关系合成：设R_1为U到V的关系，R_2为V到W的关系，则R_1与R_2的合成$R_1\circ R_2$是U到W上的一个关系，其中，$\circ$是Zadeh模糊算子，使用Min-Max模糊算子进行运算。$R_1\circ R_2$的隶属函数定义为

$$\mu_{R1}\circ R_2(u,w)=\bigvee_{v\in V}\left(\mu_{R_1}(u,v)\wedge\mu_{R_2}(v,w)\right)$$

对于有限论域，模糊关系的合成可用模糊矩阵的运算表示。从有限论域U到有限论域V的模糊关系用模糊矩阵表示为$\boldsymbol{Q}=\left(q_{ij}\right)m\times l$，从有限论域$V$到有限论域$W$的模糊关系表示为$\boldsymbol{R}=\left(r_{ij}\right)l\times n$，则对$\boldsymbol{Q}$和$\boldsymbol{R}$进行合成的结果$\boldsymbol{S}=\boldsymbol{Q}\circ\boldsymbol{R}=\left(s_{ij}\right)m\times n$为从$U$到$W$的一个模糊关系，其中，$s_{ij}=\left(q_{i1}\wedge r_{1j}\right)\vee\left(q_{i2}\wedge r_{2j}\right)\vee\cdots\vee\left(q_{in}\wedge r_{nj}\right)$，这种“小中取大”的决策原则在于给各种状态的最小可能性规定上限，体现决策者从最坏的角度进行决策。例如，若$\boldsymbol{Q}$表示U到V的弟兄关系，$\boldsymbol{R}$表示V到W的父子关系，则$\boldsymbol{S}$表示U到W的叔侄关系，如图2所示。

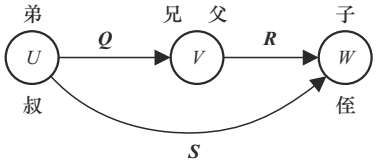

图2　模糊关系合成示意

2.4　模糊影响图

模糊影响图是模糊集合理论和贝叶斯理论对影响图理论的延伸，其中存在3类模糊集：状态模糊集、频率模糊集和模糊关系。其中，状态模糊集用模糊语言变量描述每一个节点可能的状态；频率模糊集用模糊语言变量描述每一个状态出现的频率；模糊关系用于描述每一个节点与其紧后节点的可能状态间的转移关系。

3　差分隐私算法保护效果的模糊影响图评估

3.1　方法概述

基于模糊影响图的差分隐私算法保护效果评估方法流程如图3所示，详细介绍如下。

首先，根据需要保护的隐私信息内容及格式等特征、隐私保护需求、所采用的差分隐私算法、所使用的设备等信息抽象形成场景描述，据此选取合适的评估指标，确定指标体系，并分析其影响关

系，构建差分隐私算法保护效果的模糊影响图；同时，根据每个节点的物理含义和特点，构建尽可能普适的用于描述节点状态的状态模糊集和用于描述状态发生频率的频率模糊集。其次，使用公式或调查问卷等方式度量各指标的精确值，通过模糊化环节处理得到独立节点频率矩阵。再次，通过矩阵运算，得到各层非独立节点频率矩阵，最终得到待评估的价值节点即差分隐私保护效果的概率分布，据此计算出定量的保护效果分数和定性的保护效果等级。最后，若评估结果符合预期，则输出评估结果并应用该差分隐私算法；否则，返回更换场景描述、调整控制策略或重新定义隐私操作，并对调整后的隐私保护算法重新进行评估，直至评估结果符合要求，实现迭代评估。

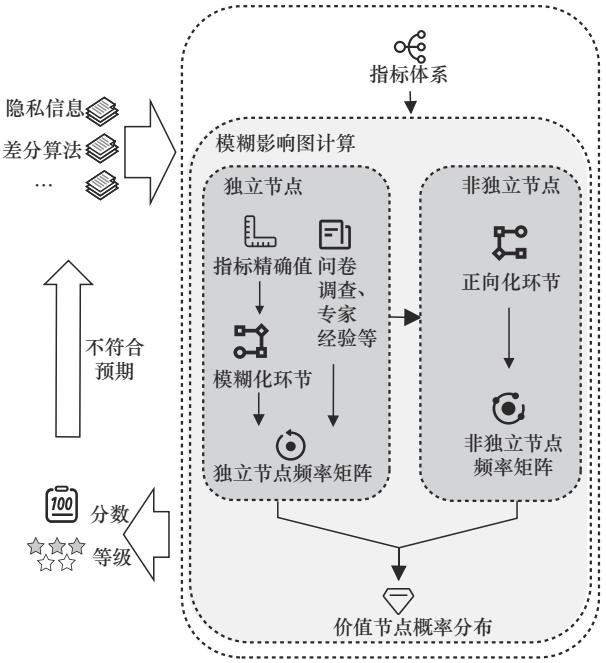

图3　基于模糊影响图的差分隐私算法保护效果评估方法流程

3.2　评估指标体系构建

以差分隐私算法的保护效果作为最终评价目标，经过广泛调研，本文分别从算法本身的角度、经算法保护后的数据特征角度和用户角度出发，将影响差分隐私算法保护效果的主要影响因素归纳总结为5个方面：算法安全性、算法可行性、隐私偏差性、数据可用性以及用户体验。下面对差分隐私算法保护效果的主要影响因素进行分析。

算法安全性是指算法能为数据提供的隐私程度。全局敏感度Δ越大、松弛项δ越小、隐私预算ε越小、隐私预算下界越紧密，算法能提供的隐私性越强。抗攻击能力是指经隐私保护后的数据抵御隐私推断的能力，可通过隐私攻击的成功率衡量，也可通过算法作用前后的信息损失衡量。

算法可行性从复杂性和可扩展性2个方面进行描述，其中，算法复杂性包括通信和计算产生的开销，可扩展性包括对不同场景和不同数据的适应性。场景适应性是指算法在新的隐私保护场景下的可移植性，数据适应性是指算法能保护的数据类型是否多样、数据规模上限以及对数据分布的要求高低。

隐私偏差性是指算法执行前后，攻击者或第三方可观测的隐私信息分量之间的偏差[1]，由噪声带来的随机性产生，该噪声可用峰度和方差描述。全局敏感度Δ、松弛项δ、隐私预算ε同样会影响隐私偏差性。

关注隐私性的同时，也应关注数据经保护后的可用性，体现在信息损失性和查询准确性。信息损失性指信息被算法作用后，对信息拥有者来说缺失了一部分可用性[1]。互信息和香农熵[30]由于能够从本质上考虑到数据的先验知识并清晰描述，被广泛应用于隐私信息量的度量。相对互信息是互信息归一化的结果，差分熵用于计算连续随机变量的信息量，根据实际情况决定是否选用二者。通过对比隐私保护前后数据在上述几个指标上的表现，可以表征信息损失性。同理，比较经隐私保护前后查询结果的误差、距离和相似度等指标，可以表征查询准确性。

考虑到隐私具有与主体主观性紧密相关的特点，将用户体验同样纳入隐私保护效果的评估体系，同时受用户对保护算法的接受程度、算法使用时的便捷程度以及用户自身的隐私保护需求影响。用户提出的隐私保护需求与其需要保护的数据紧密相关。若数据对用户来说更敏感，则数据内部关联性更大，隐私需求程度更高。

根据上述差分隐私算法保护效果影响因素的分析，按照模糊影响图的构建原则，设定差分隐私算法的保护效果作为价值节点，将各指标按照影响关系依次延伸展开，直至算法参数使用公式计算或通过问卷调查等可以直接测量的指标，例如皮尔逊相似度可以用公式计算，用户接受程度则可以通过问

卷调查得到。最终得到差分隐私算法保护效果的模糊影响图，如图4所示。

在实际应用中，受到算力、人力等资源、数据形式的限制，可能出现该图中部分独立节点指标的计算方法不适用，或独立节点过多导致效率不高、时间开销大、实用性较差的情况，需要根据使用者的需求或场景需要对影响图的结构进行调整。具体来说，可以删除现有的独立节点、增加该场景特有的新的指标或者将现有指标的计算方法进行改动，以实现对新场景的适配。有些改动会使评估结果更贴合实际，例如将距离、相似度等计算函数替换为更能反映实际情况的计算方法；有些改动可能会造成评估准确性降低、评估不全面的问题，例如将某些节点直接删减。

3.3　模糊影响图计算

模糊影响图计算的核心思想来源于模糊关系合成。对于非独立节点，前节点频率矩阵表示其状态与其频率之间的关系，前节点与后节点间的模糊关系矩阵表示前节点状态与后节点状态之间的关系。通过对上述2个矩阵进行合成，可以得到后节点频率矩阵，表示前节点频率与后节点状态之间的关系，如图5所示。

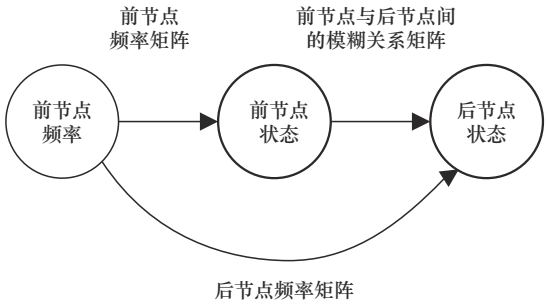

图5　非独立节点频率矩阵构造示意

对于独立节点，可使用状态向量与频率向量求取频率矩阵。按照上述方法层层传递，最终建立独立节点模糊频率与价值节点状态之间的关系。模糊影响图的计算流程如图6所示。

接下来，详细介绍模糊影响图评价算法的流程。

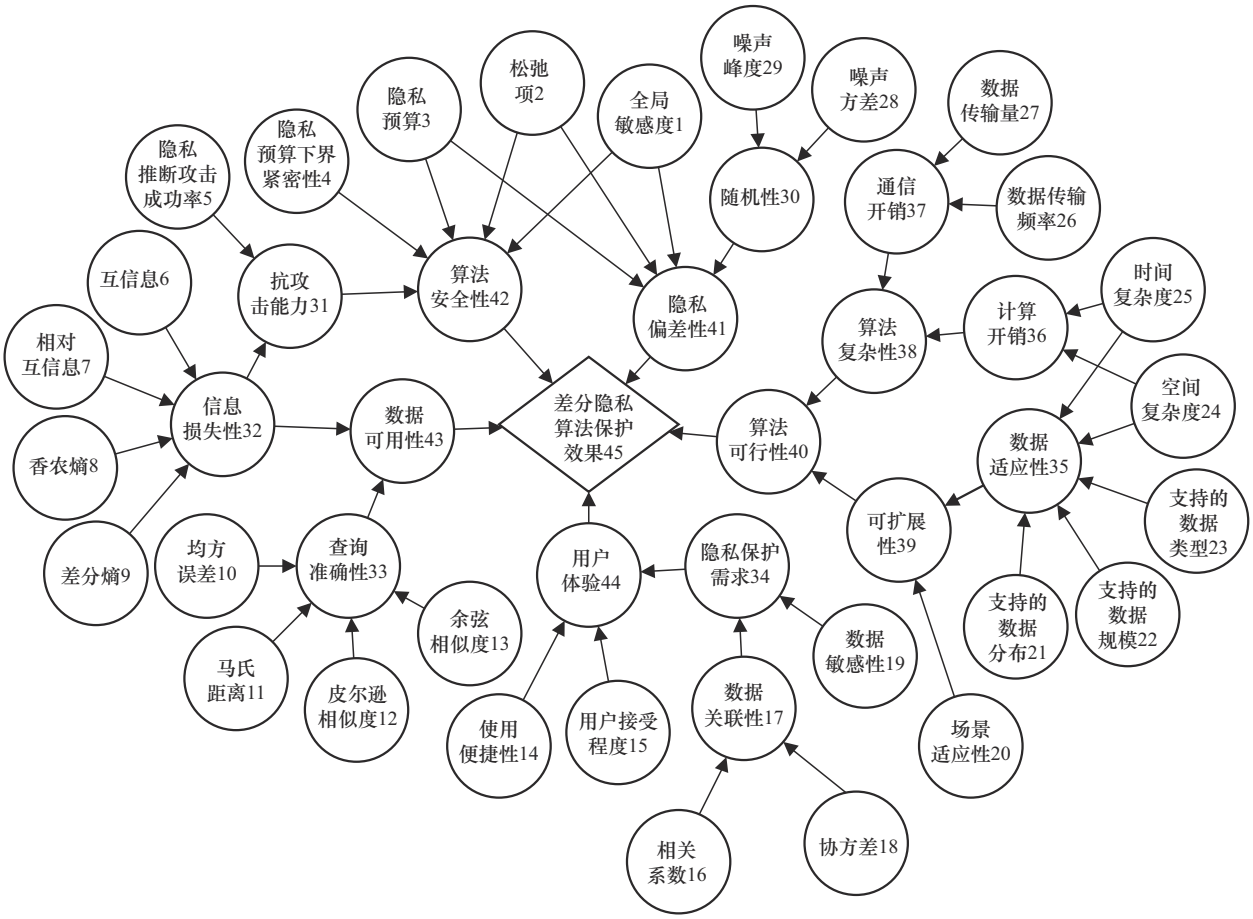

图4　差分隐私算法保护效果的模糊影响图

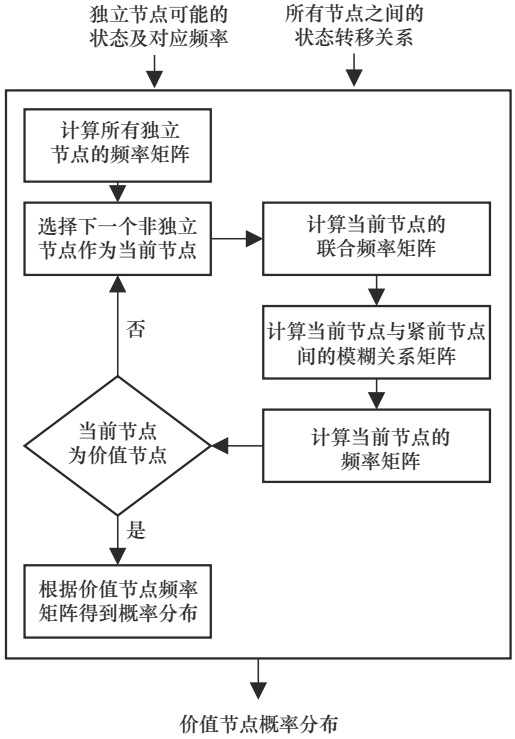

图6　模糊影响图的计算流程

1) 独立节点

通过频率模糊集和状态模糊集对独立节点进行数值结构的描述。令X表示独立节点，其所有可能的状态用模糊语言变量$P_{X_1},P_{X_2},\cdots,P_{X_n}$描述，则节点$X$的可能状态向量为

$$\boldsymbol{P}_X=\left(P_{X_1},P_{X_2},\cdots,P_{X_n}\right)^{\mathrm{T}}$$

类似地，独立节点X的频率向量为

$$\boldsymbol{f}_X=\left(f_{X_1},f_{X_2},\cdots,f_{X_n}\right)^{\mathrm{T}}$$

其中，$f_{X_1},f_{X_2},\cdots,f_{X_n}$都是模糊语言变量，用于描述节点$X$每个可能的状态相对应的频率。

在设置每个独立节点的状态与频率对应关系时，对于可以通过公式测量得到精确值的节点，根据具体含义，用相应公式计算出精确值，再经过模糊化环节（见3.4节）得到其状态模糊集和频率模糊集；对于其余不能通过公式计算、不易用数值量化的节点，根据差分隐私算法的实际情况，通过问卷调查或专家经验确定最符合的模糊语言描述，得到节点的初始状态。

独立节点X由n个可能的状态模糊集及其对应的频率模糊集组成，其频率矩阵可表示为

$$\boldsymbol{F}_X=\left(f_{X_1}\times P_{X_1}\right)\cup\left(f_{X_2}\times P_{X_2}\right)\cup\cdots\cup\left(f_{X_n}\times P_{X_n}\right)\tag{1}$$

2) 非独立节点

非独立节点的模糊关系为每个紧前节点与该非独立节点间模糊关系的并集，联合频率矩阵为所有前节点频率矩阵的并集。假设非独立节点Y有w个紧前节点$X_1,X_2,\cdots,X_w$，则节点X_t到节点Y之间的模糊关系$\boldsymbol{R}_{YX_t}$定义为

$$\begin{aligned}\boldsymbol{R}_{YX_t}=&\left(P_{X_{t_1}}\times P_{Y_1}\right)\cup\left(P_{X_{t2}}\times P_{Y_2}\right)\\&\cup\cdots\cup\left(P_{X_{t_n}}\times P_{Y_n}\right)\end{aligned}\tag{2}$$

其中，$P_{Xt_i}\in\left\{P_{X_{t_1}},P_{X_{t_2}},\cdots,P_{X_{t_n}}\right\}=P_{X_t}$，同理，$P_{Y_j}\in\left\{P_{Y_1},P_{Y_2},\cdots,P_{Y_n}\right\}=P_Y$。

为了不失一般性，假设前k个节点$X_1,X_2,\cdots,X_k,0\leqslant k\leqslant w$需要正向化（见3.4节），则节点$Y$模糊关系矩阵定义为其与所有紧前节点之间模糊关系的并集$\boldsymbol{R}_{YP}$，即

$$\begin{aligned}\boldsymbol{R}_{YP}=&\tilde{\boldsymbol{R}}_{YX_1}\cup\tilde{\boldsymbol{R}}_{YX_2}\cup\cdots\cup\tilde{\boldsymbol{R}}_{YX_k}\cup\\&\boldsymbol{R}_{YX_{k+1}}\cup\cdots\cup\boldsymbol{R}_{YX_m},0\leqslant k\leqslant w\end{aligned}\tag{3}$$

其中，$\tilde{\boldsymbol{R}}_{XY_t}$代表矩阵$\boldsymbol{R}_{YX_t}$进行左右翻转操作后的结果，若$\boldsymbol{R}_{YX_t}=\left(r_{ij}\right)h\times n$，则$\tilde{\boldsymbol{R}}_{YX_t}=\left(r'_{ij}\right)h\times n$, $r'_{ij}=r_{k(n+1-j)}$。

节点Y由所有紧前节点构成的联合频率矩阵$\boldsymbol{F}_{YP}$为

$$\begin{aligned}\boldsymbol{F}_{YP}=&\tilde{\boldsymbol{F}}_{YX_1}\cup\tilde{\boldsymbol{F}}_{YX2}\cup\cdots\cup\tilde{\boldsymbol{F}}_{YX_k}\cup\\&\boldsymbol{F}_{YX_k}+1\cup\cdots\cup\boldsymbol{F}_{YX_m},0\leqslant k\leqslant w\end{aligned}\tag{4}$$

其中，$\tilde{\boldsymbol{F}}_{YX_t}$代表矩阵$\boldsymbol{F}_{YX_t}$进行左右翻转操作后的结果。

非独立节点Y的频率矩阵通过将频率矩阵$\boldsymbol{F}_{YP}$和模糊关系矩阵$\boldsymbol{R}_{YP}$进行合成得到

$$\boldsymbol{F}_Y=\boldsymbol{F}_{YP}\circ\boldsymbol{R}_{YP}\tag{5}$$

3) 价值节点概率分布

利用上述方法计算每个节点的频率矩阵，直至获得价值节点“差分隐私算法保护效果”的频率矩阵。将其各行之和与对应频率进行乘积运算，取乘积最大的一行作为各随机结果的隶属度，从而得到价值节点随机结果的模糊集的概率函数为

$$P(x_i)=\frac{\mu_{x_i}}{\sum\limits_{\Omega_X}\mu_{x_i}} \tag{6}$$

其中，x_i表示价值节点论域中第i个元素；μ_{x_i}表示该状态下的隶属度；$\sum\limits_{\Omega_X}\mu_{x_i}$表示元素$i$所在行的隶属度值的代数和。

3.4　关键环节

1) 模糊化环节

对于算法参数和使用公式计算得到的指标，设置模糊化环节以得到各独立节点的状态值。

在该环节中，首先将各指标值归一化到[0,1]，再均匀映射到[0,100]，据此确定各个模糊状态对应的模糊频率。具体地，为了确定每个模糊状态对应的模糊频率，对于每个模糊状态，对相邻2个元素的隶属度进行线性插值，从而确定指标值对状态的隶属度，该隶属度越大则对应的频率越高，据此确定所属模糊频率。

以余弦相似度指标为例，其值越接近0，则经隐私保护前后的2列数据相关性越弱，说明数据可用性越低，因此对其进行取绝对值操作以实现归一化，再将其映射为百分制，若百分制得分对某种状态的隶属度为0.5，根据预先设定的模糊频率集，该状态对应的模糊频率为中等。

2) 正向化环节

若不对任何节点进行正向化操作，在设置独立节点的状态与频率对应关系以及各节点之间的状态映射关系时，会导致后续评估结果区分度差的问题。以图7的模糊影响图为例给出区分度问题示意，有3个节点（A、B、C），其中，独立节点A与独立节点B同时影响非独立节点C。

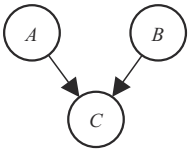

图7　区分度问题示意

在评估2个不同差分隐私算法（设为算法1和算法2）时，对于算法1，为简洁起见，假设节点A的频率矩阵为$\boldsymbol{F}_A=f_{A_1}\times P_{A_1}$，节点$B$的频率矩阵为$\boldsymbol{F}_B=f_{B_1}\times P_{B_1}$，则节点$C$的联合频率矩阵为

$$\boldsymbol{F}_{CP}=\boldsymbol{F}_A\cup\boldsymbol{F}_B=\left(f_{A_1}\times P_{A_1}\right)\cup\left(f_{B_1}\times P_{B_1}\right)$$

同理，对于算法2，假设节点A的频率矩阵为$\bar{\boldsymbol{F}}_A=\bar{f}_{A_1}\times\bar{P}_{A_1}$，节点$B$的频率矩阵为$\bar{\boldsymbol{F}}_B=\bar{f}_{B_1}\times\bar{P}_{B_1}$，则节点$C$的联合频率矩阵为

$$\bar{\boldsymbol{F}}_{CP}=\bar{\boldsymbol{F}}_A\cup\bar{\boldsymbol{F}}_B=\left(\bar{f}_{A_1}\times\bar{P}_{A_1}\right)\cup\left(\bar{f}_{B_1}\times\bar{P}_{B_1}\right)$$

若2个算法的A、B的初始化结果对称，即$(P_{A_1},P_{B_1})=(\bar{P}_{A_1},\bar{P}_{B_1})$，且$f_{A_1}=f_{B_1}=\bar{f}_{A_1}=\bar{f}_{B_1}=f$，则节点$C$的联合频率矩阵为

$$\begin{aligned}&\bar{\boldsymbol{F}}_{CP}=\left(\bar{f}_{A_1}\times\bar{P}_{A_1}\right)\cup\left(\bar{f}_{B_1}\times\bar{P}_{B_1}\right)=\\&\left(f\times\bar{P}_{A_1}\right)\cup\left(f\times\bar{P}_{B_1}\right)=\\&\left(f\times P_{B_1}\right)\cup\left(f\times P_{A_1}\right)=\boldsymbol{F}_{CP}\end{aligned}$$

虽然2个算法的初始化情况不同，但评估结果完全相同，导致该影响图模型对这2个算法的区分度较差。推广到独立节点的频率矩阵由多项组成时，只要节点A、B的初始化结果出现对称，上述问题就会发生。极端情况下，如果$\boldsymbol{F}_A$与$\boldsymbol{F}_B$的含义均表示算法1的效果非常好，$\bar{\boldsymbol{F}}_A$与$\bar{\boldsymbol{F}}_B$的含义均表示算法2的效果非常差，则此时二者的评估结果完全一致，与其含义相矛盾。

这一问题在非独立节点上也同样存在。当多个兄弟节点对下一级节点的影响趋势相反时，如信息损失值越小，数据可用性越好，而查询准确性越大，数据可用性越好，就存在出现该问题的隐患。

为解决该问题，本文引入正向化环节，在计算过程中，将同一节点的所有紧前节点对其的影响方向统一，同时需要将部分紧前节点的频率矩阵翻转。具体来说，当节点X与其多数兄弟节点对紧后节点Y的影响关系趋势不一致时，将表示该影响关系的模糊关系矩阵$\boldsymbol{R}_{YX}$进行左右翻转处理得到矩阵$\tilde{\boldsymbol{R}}_{YX}$；同时，将节点$X$的频率矩阵$\boldsymbol{F}_X$左右翻转得到矩阵$\tilde{\boldsymbol{F}}_X$，表示将该节点的含义取反，以保证计算结果的含义不变。

此外，当节点X指向多个紧后节点$Y_1,Y_2,\cdots,Y_k,Y_{k+1},\cdots,Y_w$时，不失一般性地，假设需要对前$k$个后继节点的关系进行翻转，则只有在计算$Y_1,Y_2,\cdots,Y_k$的联合频率矩阵$\boldsymbol{F}_{Y_1P},\boldsymbol{F}_{Y_2P},\cdots,\boldsymbol{F}_{Y_kP}$时，使用左右翻转后的$X$的频率矩阵$\tilde{\boldsymbol{F}}_X$参与运算，其他情况下仍直接使用$X$的频率矩阵$\boldsymbol{F}_X$参与运算。

3.5 评估结果计算

对上述结果进行分析计算，分别得出对差分隐私算法保护效果定性和定量的评估结果，即用数值表示的评估分数和用模糊语言表示的评估等级。其中，评估分数计算式为

$$\text{Score}_V = \sum_i x_i P(x_i) \tag{7}$$

评估等级根据价值节点属于论域各状态模糊集的概率，依据最大隶属原则得到。同时，计算价值节点处于各状态的隶属度的方差，以量化评估结果的可信度，计算式为

$$\sigma^2{}_V = \frac{\sum_{i=1}^{n}(\mu(x_i)-\bar{\mu})^2}{n} \tag{8}$$

其中，$\bar{\mu} = \frac{1}{n}\sum_{i=1}^{n}\mu(x_i)$表示处于各状态的隶属度的均值。由于方差可以衡量价值节点概率值的分散程度，方差越大，意味着这组概率值的变异程度越高，则该差分隐私算法的保护效果处于各状态的概率越不均衡，那么根据最大隶属原则得出的评估等级越准确，因此用方差代表评估结果的可信度。

3.6 评估结果反馈与迭代

将对该差分隐私算法保护效果的评估结果向前反馈给场景描述、隐私控制、隐私操作等环节，若评估结果未达到预期标准，则根据评估结果、用户需求、关注侧重点，选择更换场景描述（即更换被保护数据的表示形式或格式）、调整隐私控制策略（即重新选取保护算法）或重新定义隐私操作（即调整该差分隐私算法参数），对修正后的差分隐私算法进行新一轮评估，直至隐私保护效果达到预期标准；否则，反馈迭代流程结束，输出最终评估结果。

例如，当评估结果不符合预期时，考虑到修改成本、隐私需求侧重点等因素，用户可以首先选择调整算法参数，如全局敏感度Δ、松弛项δ、隐私预算ε等，并对保护效果重新进行评估，当效果不满意时，可以选择更换差分隐私算法，并对保护效果再次进行评估，若效果仍然不满意，则将被保护数据更改格式或将隐私需求更改表现形式，并对保护效果再次进行评估，直至效果符合预期，则接受该差分隐私保护方案并输出对其的评价结果。

4 实验及分析

为了验证基于模糊影响图的差分隐私算法保护效果评估方法的效果，本文借助山东电网电–碳分析数据集对本文方法予以验证。该数据集包含用电量、清洁能源占比、GDP等数据，涉及七大行业及其下属41类子行业电能关联分析、分行业分区域碳排放计算等业务。用电量数据涉及用户个人居家时段、生活规律等个人敏感信息，因此对用电量数据列使用Laplace差分隐私算法（$\varepsilon = 0.1$）进行保护，使用本文方法进行评估。

4.1 定义模糊语言变量

1) 状态模糊集

在差分隐私算法保护能力评估过程中，节点所代表的意义不同，其状态空间也不尽相同，考虑模型建立的普适性和方便性，以及评估过程的可计算性，应尽可能对所有节点构建相同的状态模糊集。

为了描述各个指标的状态，用百分制来表示其程度。将0~100分划分为11个值，即$\{0,10,20,30,40,50,60,70,80,90,100\}$，构成状态模糊集的论域，以此确定节点的状态模糊集。模糊状态划分得越细，评价越准确，但是心理学研究表明，人的最佳区分能力为6个等级左右，为保证对称性，将模糊状态划分为5个。根据对不同状态的隶属度，定义状态模糊集如下。

$$\text{程度非常高(VHL)}=\{70|0.1,80|0.3,90|0.7,100|1.0\}$$
$$\text{程度很高(HL)}=\{50|0.2,60|0.6,70|1.0,80|0.8\}$$
$$\text{程度中等(ML)}=\{30|0.2,40|0.7,50|1.0,60|0.7,70|0.2\}$$
$$\text{程度很低(LL)}=\{20|0.8,30|1.0,40|0.6,50|0.2\}$$
$$\text{程度非常低(VLL)}=\{0|1.0,10|0.7,20|0.3,30|0.1\}$$

2) 频率模糊集

节点的不同状态都以[0,1]的概率出现，然而，由于难以得到统计上要求的样本，概率分布往往根据推测得到或来自专家估计。为了方便后续计算，将其进行离散化，即将频率区间[0,1]划分为11个值，分别为$\{0,0.1,0.2,0.3,0.4,0.5,0.6,0.7,0.8,0.9,1.0\}$。使用德尔菲法，得出“高”“中”“低”的频率模糊集合；“非常高”“非常低”所对应的频率分别与“高”“低”相同，对相应频率的隶属度分别等于后者的平方。根据不同频率的隶属度，定义频率模糊集如下。

$$\text{高(H)} = \{0.7|0.5,0.8|0.7,0.9|0.9,1.0|1.0\}$$

$$\text{中(M)} = \{0.3|0.2,0.4|0.8,0.5|1.0,0.6|0.8,0.7|0.2\}$$

$$\text{低(L)} = \{0|1.0,0.1|0.9,0.2|0.7,0.3|0.5\}$$

$$\text{非常高(VH)} = \{0.7|0.25,0.8|0.49,0.9|0.81,1.0|1.0\}$$

$$\text{非常低(VL)} = \{0|1.0,0.1|0.81,0.2|0.49,0.3|0.25\}$$

4.2　确定节点状态及关系

图 4 确定的差分隐私算法保护效果的模糊影响图中共有 28 个独立节点，根据电-碳分析模型经差分隐私算法保护前后的各个性质变化以及该差分隐私算法本身的性质，对独立节点的状态模糊集和频率模糊集进行初始化，构造出状态-频率数据。

以独立节点 7“相对互信息”为例，计算应用差分隐私算法前后的用电量数据列的相对互信息为

$$\text{RelativeMutualInformation} = \frac{I(\text{OriginalData};\ \text{NoisyData})}{I(\text{OriginalData};\ \text{OriginalData})} = \frac{1.766}{5.517} = 0.320$$

其中，$I(\text{Original Data};\ \text{Noisy Data})$表示使用差分隐私算法前后的用电量数据列的互信息；$I(\text{Original Data};\ \text{Original Data})$表示用电量数据列和自身的互信息。其中，互信息的计算式为

$$I(X;\ Y) = \sum_{x \in X} \sum_{y \in Y} P(x,y) \log\left(\frac{P(x,y)}{P(x)P(y)}\right)$$

对上述结果进行模糊化，确定节点 7 处于每个可能状态的频率。上述相对互信息小于 1 且值较小，说明差分隐私算法对用电量数据列引入了一些额外的不确定性，使得变量之间的关系发生了变化，用电量数据列关系变得更加复杂或难以捕捉，因此其处于状态“VLL”的频率为“VL”，处于状态“LL”的频率为“M”，处于状态“ML”的频率为“H”。

按上述方法确定所有独立节点的初始状态，如表 1 所示。随后从独立节点开始，确定各节点间的模糊对应关系，直到价值节点，如表 2 所示，其中，→表示前节点状态向其紧后的非独立节点状态的映射，节点 45 为价值节点。

4.3　模糊影响图计算

1) 计算独立节点的频率矩阵

以独立节点 7“相对互信息”为例，根据表 1 确定的状态-频率初始化数据、节点 7 的模糊状态及相应的频率数值和式(1)，首先计算各个中间结果，即

表 1　独立节点模糊状态和模糊频率对应关系

节点	名称	状态发生的频率				
		VHL	HL	ML	LL	VLL
1	全局敏感度	—	VL	VH	—	—
2	松弛项	—	—	—	—	VH
3	隐私预算	—	M	VH	—	—
4	隐私预算下界紧密性	VH	—	—	—	—
5	隐私推断攻击成功率	VH	L	—	—	—
6	互信息	—	—	H	M	VL
7	相对互信息	—	—	H	M	VL
8	熵	—	L	H	—	—
9	差分熵	—	L	H	—	—
10	均方误差	—	VH	M	—	—
11	马氏距离	—	M	H	—	—
12	皮尔逊相似度	—	M	VH	—	—
13	余弦相似度	—	M	VH	—	—
14	使用便捷性	—	M	H	—	—
15	用户接受程度	—	L	M	VH	H
16	相关系数	—	—	VH	L	—
18	协方差	—	VL	H	—	—
19	数据敏感性	—	M	H	—	—
20	场景适应性	L	M	VH	—	—
21	支持的数据分布	L	M	H	—	—
22	支持的数据规模	L	M	H	—	—
23	支持的数据类型	L	M	H	—	—
24	空间复杂度	—	VL	H	—	—
25	时间复杂度	—	VL	H	—	—
26	数据传输频率	—	VL	VH	L	—
27	数据传输量	—	M	VH	—	—
28	噪声方差	—	—	L	VH	—
29	噪声峰度	—	VH	L	—	—

$$\text{非常低(VL)} \times \text{程度非常低(VLL)} = \begin{bmatrix} & 0 & 10 & 20 & 30 & 40 & \cdots & 100 \\ 0 & 1 & 0.7 & 0.3 & 0.1 & 0 & \cdots & 0 \\ 0.1 & 0.81 & 0.567 & 0.243 & 0.081 & 0 & \cdots & 0 \\ 0.2 & 0.49 & 0.343 & 0.147 & 0.49 & 0 & \cdots & 0 \\ 0.3 & 0.25 & 0.175 & 0.075 & 0.025 & 0 & \cdots & 0 \\ 0.4 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1.0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 \end{bmatrix},$$

再对多个中间结果进行模糊集的并运算，得到节点 7 的频率矩阵为

表2 **节点模糊状态关系**

节点	名称	紧后节点	名称	节点关系
		42	算法安全性	VHL→VHL;HL→HL;ML→ML;LL→LL;VLL→VLL
1	全局敏感度	41	隐私偏差性	VHL→VHL;HL→HL;ML→ML;LL→LL;VLL→VLL
		43	数据可用性	VHL→VLL;HL→LL;ML→ML;LL→HL;VLL→VHL
		42	算法安全性	VHL→VLL;HL→LL;ML→ML;LL→HL;VLL→VHL
2	松弛项	41	隐私偏差性	VHL→VHL;HL→HL;ML→ML;LL→LL;VLL→VLL
		43	数据可用性	VHL→VHL;HL→HL;ML→ML;LL→LL;VLL→VLL
		42	算法安全性	VHL→VHL;HL→HL;ML→ML;LL→LL;VLL→VLL
3	隐私预算	41	隐私偏差性	VHL→VHL;HL→HL;ML→ML;LL→LL;VLL→VLL
		43	数据可用性	VHL→VLL;HL→LL;ML→ML;LL→HL;VLL→VHL
4	隐私预算下界紧密性	42	算法安全性	VHL→VLL;HL→LL;ML→ML;LL→HL;VLL→VHL
5	隐私推断攻击成功率	31	抗攻击能力	VHL→VHL;HL→HL;ML→ML;LL→LL;VLL→VLL
6	互信息	32	信息损失性	VHL→VHL;HL→HL;ML→ML;LL→LL;VLL→VLL
7	相对互信息	32	信息损失性	VHL→VLL;HL→LL;ML→ML;LL→HL;VLL→VHL
8	熵	32	信息损失性	VHL→VLL;HL→LL;ML→ML;LL→HL;VLL→VHL
9	差分熵	32	信息损失性	VHL→VHL;HL→HL;ML→ML;LL→LL;VLL→VLL
10	均方误差	33	查询准确性	VHL→VHL;HL→HL;ML→ML;LL→LL;VLL→VLL
11	马氏距离	33	查询准确性	VHL→VHL;HL→HL;ML→ML;LL→LL;VLL→VLL
12	皮尔逊相似度	33	查询准确性	VHL→VHL;HL→HL;ML→ML;LL→LL;VLL→VLL
13	余弦相似度	33	查询准确性	VHL→VHL;HL→HL;ML→ML;LL→LL;VLL→VLL
14	使用便捷性	44	用户体验	VHL→VHL;HL→HL;ML→ML;LL→LL;VLL→VLL
15	用户接受程度	44	用户体验	VHL→VHL;HL→HL;ML→ML;LL→LL;VLL→VLL
16	相关系数	17	数据关联性	VHL→VHL;HL→HL;ML→ML;LL→LL;VLL→VLL
17	数据关联性	34	隐私保护需求	VHL→VHL;HL→HL;ML→ML;LL→LL;VLL→VLL
18	协方差	17	数据关联性	VHL→VHL;HL→HL;ML→ML;LL→LL;VLL→VLL
19	数据敏感性	34	隐私保护需求	VHL→VHL;HL→HL;ML→ML;LL→LL;VLL→VLL
20	场景适应性	39	可扩展性	VHL→VHL;HL→HL;ML→ML;LL→LL;VLL→VLL
21	支持的数据分布	35	数据适应性	VHL→VLL;HL→LL;ML→ML;LL→HL;VLL→VHL
22	支持的数据规模	35	数据适应性	VHL→VHL;HL→HL;ML→ML;LL→LL;VLL→VLL
23	支持的数据类型	35	数据适应性	VHL→VLL;HL→LL;ML→ML;LL→HL;VLL→VHL
24	空间复杂度	35	数据适应性	VHL→VHL;HL→HL;ML→ML;LL→LL;VLL→VLL
		36	计算开销	VHL→VHL;HL→HL;ML→ML;LL→LL;VLL→VLL
25	时间复杂度	35	数据适应性	VHL→VHL;HL→HL;ML→ML;LL→LL;VLL→VLL
		36	计算开销	VHL→VHL;HL→HL;ML→ML;LL→LL;VLL→VLL
26	数据传输频率	37	通信开销	VHL→VLL;HL→LL;ML→ML;LL→HL;VLL→VHL
27	数据传输量	37	通信开销	VHL→VHL;HL→HL;ML→ML;LL→LL;VLL→VLL
28	噪声方差	30	算法随机性	VHL→VHL;HL→HL;ML→ML;LL→LL;VLL→VLL
29	噪声峰度	30	算法随机性	VHL→VHL;HL→HL;ML→ML;LL→LL;VLL→VLL
30	算法随机性	41	隐私偏差性	VHL→VLL;HL→LL;ML→ML;LL→HL;VLL→VHL
31	抗攻击能力	42	算法安全性	VHL→VHL;HL→HL;ML→ML;LL→LL;VLL→VLL
32	信息损失性	31	抗攻击能力	VHL→VHL;HL→HL;ML→ML;LL→LL;VLL→VLL
		43	数据可用性	VHL→VHL;HL→HL;ML→ML;LL→LL;VLL→VLL
33	查询准确性	43	数据可用性	VHL→VHL;HL→HL;ML→ML;LL→LL;VLL→VLL
34	隐私保护需求	44	用户体验	VHL→VHL;HL→HL;ML→ML;LL→LL;VLL→VLL
35	数据适应性	39	可扩展性	VHL→VHL;HL→HL;ML→ML;LL→LL;VLL→VLL
36	计算开销	38	算法复杂性	VHL→VHL;HL→HL;ML→ML;LL→LL;VLL→VLL
37	通信开销	38	算法复杂性	VHL→VHL;HL→HL;ML→ML;LL→LL;VLL→VLL
38	算法复杂性	40	算法可行性	VHL→VHL;HL→HL;ML→ML;LL→LL;VLL→VLL
39	可扩展性	40	算法可行性	VHL→VHL;HL→HL;ML→ML;LL→LL;VLL→VLL
40	算法可行性	45	隐私保护算法能力	VHL→VHL;HL→HL;ML→ML;LL→LL;VLL→VLL
41	隐私偏差性	45	隐私保护算法效果	VHL→VHL;HL→HL;ML→ML;LL→LL;VLL→VLL
42	算法安全性	45	隐私保护算法效果	VHL→VHL;HL→HL;ML→ML;LL→LL;VLL→VLL
43	数据可用性	45	隐私保护算法效果	VHL→VHL;HL→HL;ML→ML;LL→LL;VLL→VLL
44	用户体验	45	隐私保护算法效果	VHL→VLL;HL→LL;ML→ML;LL→HL;VLL→VHL

$$\boldsymbol{F}_7 = \{\text{高(H)} \times \text{程度中等(ML)}\} \cup \{\text{中(M)} \times \text{程度很低(LL)}\} \cup \{\text{非常低(VL)} \times \text{程度非常低(VLL)}\} =$$

	0	10	20	30	40	50	60	70	80	90	100
0	1	0.7	0.3	0.1	0	0	0	0	0	0	0
0.1	0.81	0.567	0.243	0.081	0	0	0	0	0	0	0
0.2	0.49	0.343	0.147	0.49	0	0	0	0	0	0	0
0.3	0.25	0.175	0.16	0.2	0.12	0.04	0	0	0	0	0
0.4	0	0	0.64	0.8	0.48	0.16	0	0	0	0	0
0.5	0	0	0.8	1	0.6	0.2	0	0	0	0	0
0.6	0	0	0.64	0.8	0.48	0.16	0	0	0	0	0
0.7	0	0	0.16	0.2	0.35	0.5	0.35	0.1	0	0	0
0.8	0	0	0	0.14	0.49	0.7	0.49	0.14	0	0	0
0.9	0	0	0	0.18	0.63	0.9	0.63	0.18	0	0	0
1.0	0	0	0	0.2	0.7	1	0.7	0.2	0	0	0

$\boldsymbol{F}_7$是一个模糊矩阵，即从频率模糊集到状态模糊集一个模糊关系，表示节点7可能的状态及每种状态可能发生的频率，其中，每个元素对应一组状态和频率，其值表示事件“对应状态以对应频率发生”对该关系的隶属度，例如，(0.3,20)位置处的值0.16表示：相对互信息的程度是20、出现频率是0.3对“程度中等（ML）发生的频率高（H）”的相关程度为0，对“程度很低（LL）发生的频率中（M）”的相关程度为0.16，对“程度非常低（VLL）发生的频率非常低（VL）”的相关程度为0.075；进行并集运算后，其对模糊关系$\boldsymbol{F}_7$的相关程度为0.16。以此方法计算得到图中所有独立节点的频率矩阵。

2) 计算非独立节点的频率矩阵

以非独立节点32“信息损失性”为例，该节点的前节点为独立节点6~独立节点9。首先根据表2中的节点关系，可以得到节点6~9分别与节点32的模糊对应关系，依据式(2)和式(3)，节点32的联合关系矩阵$\boldsymbol{R}_{32P}$为

$$\boldsymbol{R}_{32P} = \boldsymbol{R}'_{6-32} \cup \boldsymbol{R}'_{7-32} \cup \boldsymbol{R}_{8-32} \cup \boldsymbol{R}_{9-32} =$$

	0	10	20	30	40	50	60	70	80	90	100
0	1	0.7	0.3	0.1	0	0	0	0	0	0	0
10	0.7	0.49	0.21	0.07	0	0	0	0	0	0	0
20	0.3	0.21	0.64	0.8	0.48	0.16	0	0	0	0	0
30	0.1	0.07	0.8	1	0.6	0.2	0.14	0.04	0	0	0
40	0	0	0.48	0.6	0.49	0.7	0.49	0.14	0	0	0
50	0	0	0.16	0.2	0.7	1	0.7	0.2	0.16	0	0
60	0	0	0	0.14	0.49	0.7	0.49	0.6	0.48	0	0
70	0	0	0	0.04	0.14	0.2	0.6	1	0.8	0.07	0.1
80	0	0	0	0	0	0.16	0.48	0.8	0.64	0.21	0.3
90	0	0	0	0	0	0	0	0.07	0.21	0.49	0.7
100	0	0	0	0	0	0	0	0.1	0.3	0.7	1

依据式(4)，由前述4个节点的频率矩阵计算节点32的联合频率矩阵$\boldsymbol{F}_{32P}$，其中节点6和节点7需要进行正向化处理，因此使用翻转后的频率矩阵参与此步运算。

$$\boldsymbol{F}_{32P}=\boldsymbol{F}'_6\cup\boldsymbol{F}'_7\cup\boldsymbol{F}_8\cup\boldsymbol{F}_9=$$

	0	10	20	30	40	50	60	70	80	90	100
0	0	0	0	0	0	0.2	0.6	1	0.8	0.7	1
0.1	0	0	0	0	0	0.18	0.54	0.9	0.72	0.567	0.81
0.2	0	0	0	0	0	0.14	0.42	0.7	0.56	0.343	0.49
0.3	0	0	0	0	0	0.1	0.3	0.5	0.4	0.175	0.25
0.4	0	0	0	0	0	0.16	0.48	0.8	0.64	0.	0
0.5	0	0	0	0	0	0.2	0.6	1	0.8	0	0
0.6	0	0	0	0	0	0.16	0.48	0.8	0.64	0	0
0.7	0	0	0	0.1	0.35	0.5	0.35	0.2	0..16	0	0
0.8	0	0	0	0.14	0.49	0.7	0.49	0.14	0	0	0
0.9	0	0	0	0.18	0.63	0.9	0.63	0.18	0	0	0
1.0	0	0	0	0.2	0.7	1	0.7	0.2	0	0	0

根据式(5)计算非独立节点32的频率矩阵$\boldsymbol{F}_{32}$为

$$\boldsymbol{F}_{32}=\boldsymbol{F}_{32P}\circ\boldsymbol{R}_{32P}=$$

	0	10	20	30	40	50	60	70	80	90	100
0	0	0	0.16	0.2	0.49	0.6	0.6	1	0.8	0.7	1
0.1	0	0	0.16	0.18	0.49	0.54	0.6	0.9	0.8	0.7	0.81
0.2	0	0	0.14	0.14	0.42	0.42	0.6	0.7	0.7	0.49	0.49
0.3	0	0	0.10	0.14	0.3	0.3	0.5	0.5	0.5	0.25	0.3
0.4	0	0	0.16	0.16	0.48	0.48	0.6	0.8	0.8	0.21	0.3
0.5	0	0	0.16	0.2	0.49	0.6	0.6	1	0.8	0.21	0.3
0.6	0	0	0.16	0.16	0.48	0.48	0.6	0.8	0.8	0.21	0.3
0.7	0.1	0.07	0.35	0.35	0.5	0.5	0.5	0.35	0.35	0.16	0.16
0.8	0.1	0.07	0.48	0.49	0.7	0.7	0.7	0.49	0.48	0.07	0.1
0.9	0.1	0.07	0.48	0.6	0.7	0.9	0.7	0.9	0.48	0.07	0.1
1.0	0.1	0.07	0.48	0.6	0.7	1	0.7	0.6	0.48	0.07	0.1

根据上述方法，从独立节点开始计算，按顺序计算至价值节点45，则算法保护能力节点的频率矩阵$\boldsymbol{F}_{45}$为

$$\boldsymbol{F}_{45}=$$

	0	10	20	30	40	50	60	70	80	90	100	SUM
0	1	0.7	0.8	1	0.7	1	0.7	1	0.8	0.7	1	9.4
0.1	0.81	0.7	0.8	0.9	0.7	0.9	0.7	0.9	0.8	0.7	0.9	8.81
0.2	0.49	0.49	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	7.28
0.3	0.3	0.3	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	5.1
0.4	0.3	0.3	0.8	0.8	0.7	0.8	0.7	0.8	0.8	0.3	0.3	6.6
0.5	0.3	0.3	0.8	1	0.7	1	0.7	1	0.8	0.3	0.3	7.2
0.6	0.3	0.3	0.8	0.8	0.7	0.8	0.7	0.8	0.8	0.3	0.3	6.6
0.7	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.3	0.3	5.1
0.8	0.7	0.7	0.6	0.6	0.7	0.7	0.7	0.6	0.6	0.49	0.49	6.88
0.9	0.9	0.7	0.8	0.81	0.7	0.9	0.7	0.6	0.6	0.7	0.81	8.22
1.0	1	0.7	0.8	1	0.7	1	0.7	0.6	0.6	0.7	1	8.8

3) 评估结果计算

根据价值节点的频率矩阵，分析得到评估结果。频率矩阵$\boldsymbol{F}_{45}$中SUM列代表矩阵每行数值的和，根据式(6)选取SUM列与频率的乘积最大的一行，即频率为1.0的行作为随机结果的隶属度，并计算其概率分布和累积概率，如表3和图8的浅色柱状图所示。

分析表3，根据式(7)可计算保护效果的分数为

$$\begin{aligned}\text{Score}_{45}=&0.1136\times0+0.0795\times10+\\&0.0909\times20+0.1136\times30+0.0795\times40+\\&0.1136\times50+0.0795\times60+0.0682\times70+\\&0.0682\times80+0.0795\times90+0.1136\times100=48.4\end{aligned}$$

同时，保护效果在0~30的概率为0.3295，20~50的概率为0.329 5，30~70的概率为0.454 5，50~

80 的概率为 0.397 7，70~100 的概率为 0.397 7，如表 4 所示。由此，根据最大隶属原则，得出所用 Laplace 机制本次对用电量进行保护的效果等级为中（ML）。

表 3　评估 Laplace 机制的价值节点概率分布

隐私保护算法效果	隶属度	概率	累积概率
0	1.0	0.113 6	0.113 6
10	0.7	0.079 5	0.193 2
20	0.8	0.090 9	0.284 1
30	1.0	0.113 6	0.397 7
40	0.7	0.079 5	0.477 3
50	1.0	0.113 6	0.590 9
60	0.7	0.079 5	0.670 5
70	0.6	0.068 2	0.738 6
80	0.6	0.068 2	0.806 8
90	0.7	0.079 5	0.886 4
100	1.0	0.113 6	1.000 0

根据式(8)计算上述评估结果的可信度为

$$\sigma^2{}_{45} = 0.028$$

4.4　结果反馈迭代

上述对 Laplace 机制的评估结果分值较低，选择重新定义隐私操作，改为使用 Gaussian 机制对用电量进行保护，即对用电量数据列使用 Gaussian 差分隐私算法（$\varepsilon = 0.1, \delta = 0.001$）进行保护，并再次进行评估。重新根据应用场景、被保护数据和算法特征等形成场景描述，仍使用图 4 所示的模糊影响图，重新度量各指标的值并对独立节点进行初始化，经过相同的计算流程，得到价值节点 45 的频率矩阵为

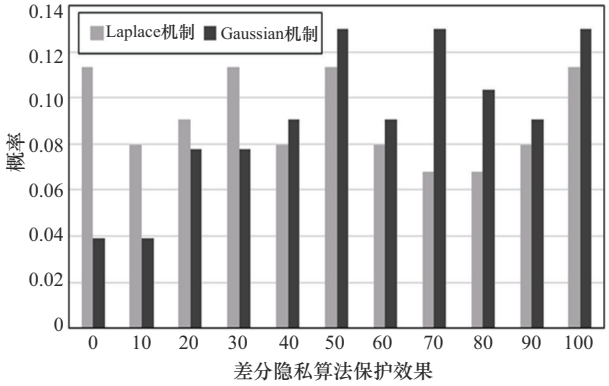

图 8　价值节点概率分布

表 4　Laplace 机制评估结果

保护效果	概率值
非常差	0.329 5
差	0.329 5
中	0.454 5
好	0.397 7
非常好	0.397 7

$$\boldsymbol{F}'_{45} = \begin{bmatrix} & 0 & 10 & 20 & 30 & 40 & 50 & 60 & 70 & 80 & 90 & 100 & \text{SUM} \\ 0 & 0.3 & 0.3 & 0.8 & 1.0 & 0.7 & 1.0 & 0.7 & 1.0 & 0.8 & 0.7 & 1.0 & 8.3 \\ 0.1 & 0.3 & 0.3 & 0.8 & 0.9 & 0.7 & 0.9 & 0.7 & 0.9 & 0.8 & 0.7 & 0.9 & 7.9 \\ 0.2 & 0.3 & 0.3 & 0.7 & 0.7 & 0.7 & 0.7 & 0.7 & 0.7 & 0.7 & 0.7 & 0.7 & 6.9 \\ 0.3 & 0.3 & 0.3 & 0.5 & 0.5 & 0.5 & 0.5 & 0.5 & 0.5 & 0.5 & 0.5 & 0.5 & 5.1 \\ 0.4 & 0.3 & 0.3 & 0.8 & 0.8 & 0.6 & 0.6 & 0.6 & 0.8 & 0.8 & 0.7 & 0.8 & 7.1 \\ 0.5 & 0.3 & 0.3 & 0.8 & 1.0 & 0.6 & 0.6 & 0.6 & 1.0 & 0.8 & 0.7 & 1.0 & 7.7 \\ 0.6 & 0.3 & 0.3 & 0.8 & 0.8 & 0.6 & 0.6 & 0.6 & 0.8 & 0.8 & 0.7 & 0.8 & 7.1 \\ 0.7 & 0.3 & 0.3 & 0.5 & 0.5 & 0.5 & 0.5 & 0.5 & 0.5 & 0.5 & 0.3 & 0.3 & 4.7 \\ 0.8 & 0.3 & 0.3 & 0.6 & 0.6 & 0.7 & 0.7 & 0.7 & 0.6 & 0.6 & 0.49 & 0.49 & 6.08 \\ 0.9 & 0.3 & 0.3 & 0.6 & 0.6 & 0.7 & 0.7 & 0.7 & 0.81 & 0.8 & 0.7 & 0.81 & 7.22 \\ 1.0 & 0.3 & 0.3 & 0.6 & 0.6 & 0.7 & 0.7 & 0.7 & 1.0 & 0.8 & 0.7 & 1.0 & 7.7 \end{bmatrix}$$

选取频率为 1.0 的行作为随机结果的隶属度，其概率分布如表 5 所示，累积概率如图 8 的深色柱状图所示。

根据表 5，计算所用 Gaussian 机制的评估分数为

$$\text{Score}'_{45} = 58.4$$

同时，如表 6 所示，根据最大隶属原则，得出所用 Gaussian 机制本次对用电量进行保护的效果等级为中（ML）。可信度为

$$\sigma'^2{}_{45} = 0.062$$

若仍不符合预期，可以多次调整迭代。

由上述结果可见，虽然二者等级相同，但综合得分略有差距，经分析，产生这种现象的原因可能

是Laplace机制作为经典的差分隐私算法，为了保证完全的隐私性，加入过多噪声，使得可用性下降较为明显，同时对用户体验有一定影响；相比之下，Gaussian机制作为具有松弛项的差分隐私算法，牺牲了小部分隐私性，在保证较高程度隐私性的同时提升了可用性，同时收获了较好的用户体验。

表5　评估Gaussian机制的价值节点概率分布

隐私保护算法效果	隶属度	概率	累积概率
0	0.3	0.039 0	0.039 0
10	0.3	0.039 0	0.077 9
20	0.6	0.077 9	0.155 8
30	0.6	0.077 9	0.233 8
40	0.7	0.090 9	0.324 7
50	1.0	0.129 9	0.454 5
60	0.7	0.090 9	0.545 5
70	1.0	0.129 9	0.675 3
80	0.8	0.103 9	0.779 2
90	0.7	0.090 9	0.870 1
100	1.0	0.129 9	1.000 0

表6　Gaussian机制评估结果

保护效果	概率值
非常差	0.454 5
差	0.454 5
中	0.515 9
好	0.376 6
非常好	0.233 8

4.5　对比实验

本节将本文方法与目前较为成熟的多层模糊综合评价方法[20]进行比较。考虑到多层模糊综合评价方法的指标体系无法处理影响关系的交叉，因此对图4所示的模糊影响图中产生交叉的影响弧进行删减，形成指标体系a，如图9所示。

由于交叉的影响弧理论上有不同的删减方式，为了减少这一处理可能带来的影响，因此将指标1、2、3、32的删减方式进行调整，形成指标体系b，并重复实验。

使用本文方法的状态模糊集作为评语集V，使用上述指标体系中的非叶子节点作为因素集U，叶子节点作为具体指标。对于受多个下一级因素影响

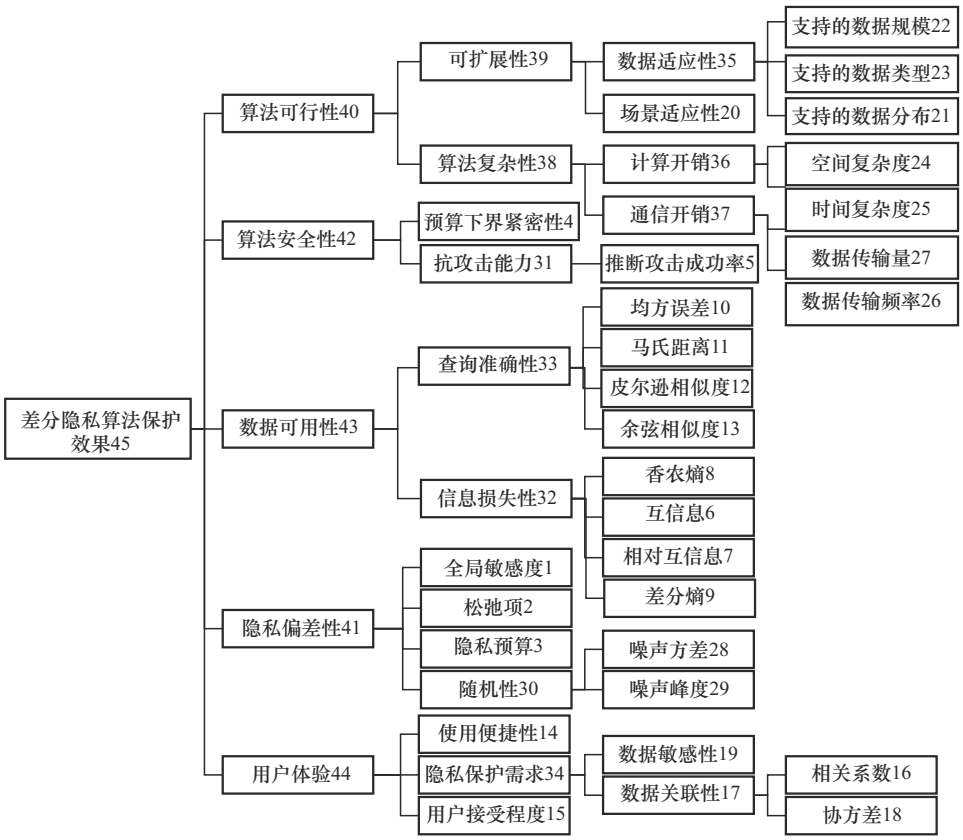

图9　指标体系a

的因素，各个下一级因素平均分配权重，据此构建各权值向量$\boldsymbol{A}$。在构造底层因素的模糊评估矩阵$\boldsymbol{R}$时，每个行向量由对应指标值对每个评语的隶属度构成，其中隶属度由4.2节中该指标对应的独立节点的初始状态转换得到。

由于形成的指标体系层数不均衡，存在因素与指标属于同一层的情况，因此需要调整二者父节点的模糊评估矩阵构造方式。具体地，以“隐私保护需求”因素为例，其模糊评估矩阵由“数据关联性”因素的行向量与“数据敏感性”指标的行向量共同构成。其中，“数据关联性”因素的行向量为其评估结果向量$\boldsymbol{B}$，而“数据敏感性”指标的行向量则直接由隶属度构成。

使用指标体系a分别就Laplace机制（$\varepsilon = 0.1$）和Gaussian机制（$\varepsilon = 0.1,\delta = 0.001$）对用电量的保护效果进行评估，得到2个算法的评估结果向量分别为$\boldsymbol{B}_{\text{a_Laplace}}$=[0.200 0,0.177 3,0.470 6,0.08166,0.070 5]和$\boldsymbol{B}_{\text{a_Gaussian}}$ = [0.190 5,0.135 8,0.551 6,0.113 6,0.008 5]，二者对等级“中”的隶属度均最高，根据最大隶属原则，对二者的评估结果均为“中”，如表7所示。

表7　　模糊综合评价结果a

差分隐私算法	保护效果
Laplace机制	中（ML）
Gaussian机制	中（ML）

使用指标体系b按上述方式再次进行评估，得到2个算法的模糊评价矩阵分别为$\boldsymbol{B}_{\text{b_Laplace}}$ = [0.0680,0.2674,0.4673,0.1372,0.0601] 和 $\boldsymbol{B}_{\text{b\_Gaussian}}$ = [0.0695,0.2063,0.5490,0.1683,0.0068]，二者的评估结果均为“中”，如表8所示。

表8　　模糊综合评价结果b

差分隐私算法	保护效果
Laplace机制	中（ML）
Gaussian机制	中（ML）

所使用的Gaussian机制相较Laplace机制而言，在数据可用性、用户体验、可扩展性等方面有较大提升，因此总体评价也应更高，而上述评估结果无法很好地反映出这一差距，由此说明本文方法具有更好的评估效果。

4.6　消融实验

在消融实验中，删除本文方法中的正向化环节，其他部分保持不变，通过对比表明该环节对评估结果的影响，进而说明正向化环节存在的必要性。

将正向化环节删除后，分别就Laplace机制（$\varepsilon = 0.1$）和Gaussian机制（$\varepsilon = 0.1,\delta = 0.001$）对用电量的保护效果进行评估，结果如表9所示。

由表9可知，此时的评估方法对2种保护算法没有区分度。这是由于删除了正向化环节，无法区分初始化时出现的对称情况，因此删除正向化环节会使该方法灵敏度降低，故该环节是不可或缺的。

5　结束语

针对现有的隐私保护效果评估指标不全面、没有考虑指标间制约关系、无法给出综合指标的问题，本文提供一种基于模糊影响图的差分隐私算法保护效果评估方法，并得出直观的、便于比较的分数和等级。该方法建立差分隐私算法保护效果评估指标体系，利用模糊理论处理评估过程的不确定性形成模糊影响图，计算得到保护效果的分数和等级，并评估结果的可信度。在计算过程中，引入正向化环节解决部分情况下评估结果区分度差的问题。最后，将评估结果反馈到场景描述、隐私控制或隐私操作过程中，实现为隐私信息提供更好的保护效果。实验结果反映了不同差分隐私算法的差异，并给出了合理解释。

在未来工作中，在节点的模糊集合确定阶段，可以将状态划分得更加细致，以获得更大的取值空间；在模糊关系传递阶段，可以引入权重，从而体现出不同因素对节点影响重要性的区别。

表9　　删除正向化环节的评估结果

差分隐私算法	保护效果的概率值					评估分数	保护效果
	非常好	好	中	差	非常差		
Laplace机制	0.372 3	0.372 3	0.468 1	0.372 3	0.372 3	50.0	中（ML）
Gaussian机制	0.372 3	0.372 3	0.468 1	0.372 3	0.372 3	50.0	中（ML）

参考文献：

[1] 李凤华, 李晖, 贾焰, 等. 隐私计算研究范畴及发展趋势[J]. 通信学报, 2016, 37(4): 1-11.
LI F H, LI H, JIA Y, et al. Privacy computing: concept, connotation and its research trend[J]. Journal on Communications, 2016, 37(4): 1-11.

[2] 张文静, 刘樵, 朱辉. 基于信息论方法的多等级位置隐私度量与保护[J]. 通信学报, 2019, 40(12): 51-59.
ZHANG W J, LIU Q, ZHU H. Evaluation and protection of multi-level location privacy based on an information theoretic approach[J]. Journal on Communications, 2019, 40(12): 51-59.

[3] ZAMANI A, OECHTERING T J, SKOGLUND M. On the privacy-utility trade-off with and without direct access to the private data[J]. IEEE Transactions on Information Theory, 2024, 70(3): 2177-2200.

[4] CUFF P, YU L Q. Differential privacy as a mutual information constraint[C]//Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security. New York: ACM Press, 2016: 43-54.

[5] JAGIELSKI M, ULLMAN J, OPREA A. Auditing differentially private machine learning: how private is private SGD?[C]//Advances in Neural Information Processing Systems. Virtual Event: NIPS Foundation, 2020: 22205-22216.

[6] WANG C X, TAY W P. On the relationship between information-theoretic privacy metrics and probabilistic information privacy[J]. arXiv Preprint, arXiv:2301.08401, 2023.

[7] ASOODEH S, DIAZ M, ALAJAJI F, et al. Privacy-aware guessing efficiency[C]//Proceedings of the 2017 IEEE International Symposium on Information Theory (ISIT). Piscataway: IEEE Press, 2017: 754-758.

[8] ASOODEH S, DIAZ M, ALAJAJI F, et al. Estimation efficiency under privacy constraints[J]. IEEE Transactions on Information Theory, 2019, 65(3): 1512-1534.

[9] RASSOULI B, GÜNDÜZ D. Optimal utility-privacy trade-off with total variation distance as a privacy measure[C]//Proceedings of the 2018 IEEE Information Theory Workshop (ITW). Piscataway: IEEE Press, 2018: 1-5.

[10] JAYARAMAN B, EVANS D E. Evaluating differentially private machine learning in practice[C]//28th USENIX Security Symposium (USENIX Security). Berkeley: USENIX Association, 2019: 1895-1912.

[11] RYU J, ZHENG Y F, GAO Y S, et al. Can differential privacy practically protect collaborative deep learning inference for IoT?[J]. Wireless Networks, 2024, 30:4713-4733.

[12] DING Z Y, WANG Y X, WANG G H, et al. Detecting violations of differential privacy[C]//Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security. New York: ACM Press, 2018: 475-489.

[13] BICHSEL B, GEHR T, DRACHSLER-COHEN D, et al. DP-finder: finding differential privacy violations by sampling and optimization[C]//Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security. New York: ACM Press, 2018: 508-524.

[14] BICHSEL B, STEFFEN S, BOGUNOVIC I, et al. DP-sniper: black-box discovery of differential privacy violations using classifiers[C]//Proceedings of the 2021 IEEE Symposium on Security and Privacy (SP). Piscataway: IEEE Press, 2021: 391-409.

[15] NIU B, ZHOU Z J, CHEN Y H, et al. DP-opt: identify high differential privacy violation by optimization[C]//International Conference on Wireless Algorithms, Systems, and Applications. Berlin: Springer, 2022: 406-416.

[16] ASKIN Ö, KUTTA T, DETTE H. Statistical quantification of differential privacy: a local approach[C]//Proceedings of the 2022 IEEE Symposium on Security and Privacy (SP). Piscataway: IEEE Press, 2022: 402-421.

[17] SAKIB S K, AMARIUCAI G T, GUAN Y. Variations and extensions of information leakage metrics with applications to privacy problems with imperfect statistical information[C]//Proceedings of the 2023 IEEE 36th Computer Security Foundations Symposium (CSF). Piscataway: IEEE Press, 2023: 407-422.

[18] YE Q Q, HU H B, MENG X F, et al. PrivKV: key-value data collection with local differential privacy[C]//Proceedings of the 2019 IEEE Symposium on Security and Privacy (SP). Piscataway: IEEE Press, 2019: 317-331.

[19] OYA S, TRONCOSO C, PÉREZ-GONZÁLEZ F. Back to the drawing board: Revisiting the design of optimal location privacy-preserving mechanisms[C]//Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security (CCS). New York: ACM Press, 2017: 1959-1972.

[20] WANG Y X, DING Z Y, KIFER D, et al. CheckDP: an automated and integrated approach for proving differential privacy or finding precise counterexamples[C]//Proceedings of the 2020 ACM SIGSAC Conference on Computer and Communications Security. New York: ACM Press, 2020: 919-938.

[21] CHEN J, WANG C H, HE K, et al. Semantics-aware privacy risk assessment using self-learning weight assignment for mobile apps[J]. IEEE Transactions on Dependable and Secure Computing, 2021, 18(1): 15-29.

[22] 俞艺涵, 付钰, 吴晓平. 基于多层模糊综合评估的隐私保护效果评估方法[J]. 网络与信息安全学报, 2020, 6(6): 121-127.
YU Y H, FU Y, WU X P. Evaluation method of privacy protection effect based on multi-layer fuzzy comprehensive evaluation[J]. Chinese Journal of Network and Information Security, 2020, 6(6): 121-127.

[23] 刘金兰, 韩文秀, 李光泉. 关于工程项目风险分析的模糊影响图方法[J]. 系统工程学报, 1994, 9(2): 81-88.
LIU J L, HAN W X, LI G Q. A fuzzy influence diagram method for analyzing engineering project risk[J]. Journal of Systems Engineering, 1994, 9(2): 81-88.

[24] 程铁信, 王平, 张伟波. 模糊影响图评价算法的探讨[J]. 系统工程学报, 2004, 19(2): 177-182.
CHENG T X, WANG P, ZHANG W B. Investigation on fuzzy influence diagrams evaluation algorithm[J]. Journal of Systems Engineering, 2004, 19(2): 177-182.

[25] 刘玉梅, 陈云, 赵聪聪, 等. 高速列车传动系可靠性的外部影响因素评估[J]. 西南交通大学学报, 2019, 54(3): 535-541.
LIU Y M, CHEN Y, ZHAO C C, et al. Assessment for external influence factors of high-speed train transmission reliability[J]. Journal of Southwest Jiaotong University, 2019, 54(3): 535-541.

[26] 门金柱, 张本辉, 姚科明, 等. 基于模糊影响图的舰载直升机作战环境评估方法[J]. 火力与指挥控制, 2022, 47(10): 46-51, 58.

MEN J Z, ZHANG B H, YAO K M, et al. Evaluation method of operational environment for shipboard helicopter based on fuzzy influence diagram[J]. Fire Control & Command Control, 2022, 47(10): 46-51, 58.

[27] XIA J Y, PI Z Y, FANG W G. Predicting war outcomes based on a fuzzy influence diagram[J]. International Journal of Fuzzy Systems, 2021, 23(4): 984-1002.

[28] HOWARD R A, MATHESON J E. Influence diagrams[J]. Decision Analysis, 2005, 2(3): 127-143.

[29] ZADEH L A. Fuzzy sets[J]. Elsevier Information and Control, 1965, 8(3): 338-353.

[30] SHANNON C E. A mathematical theory of communication[J]. The Bell System Technical Journal, 1948, 27(3): 379-423.

[作者简介]

田月池（1998– ），女，河北保定人，中国科学院信息工程研究所博士生，主要研究方向为隐私计算、隐私保护效果评估。

李凤华（1966– ），男，湖北浠水人，博士，中国科学院信息工程研究所研究员、博士生导师，主要研究方向为网络与系统安全、信息保护、隐私计算。

周泽峻（2000– ），男，河南洛阳人，中国科学院信息工程研究所博士生，主要研究方向为隐私计算、数据安全。

孙哲（1987– ），男，安徽安庆人，博士，广州大学副教授，主要研究方向为隐私计算、数据安全。

郭守坤（1994– ），男，河南周口人，中国科学院信息工程研究所工程师，主要研究方向为隐私计算、数据安全。

牛犇（1984– ），男，陕西西安人，博士，中国科学院信息工程研究所研究员、博士生导师，主要研究方向为数据安全、隐私计算。