ECTI Transactions on Computer and Information Technology
Journal homepage: https://ph01.tci-thaijo.org/index.php/ecticit/
Published by the ECTI Association, Thailand, ISSN: 2286-9131

An Effective Privacy Preservation Model for Rating Datasets based on Aggregate Query Frameworks in conjunction with $(l^{p_1}, \ldots, l^{p_n})$-Privacy Constraints

Surapon Riyana[1], Kittikorn Sasujit[2], Nigran Homdoung[3], Tanate Chaichana[4] and Tammasak Punsaensri[5]

ABSTRACT

Rating datasets are data collections of user profile tuples that generally include their person-specific data and the preferred level of their interesting artifacts. Generally, rating datasets are generated for use by recommendation methods that are available in real-life systems such as mobile applications, location-based services, e-commerce systems, and websites. Aside from recommendation methods, we can see that rating datasets can further be utilized by data analysts with appropriate business reasons such as improving system policies and constructing business reports. Such a data utilization of rating datasets can lead to privacy violation issues. To address these issues in rating datasets, $(l^{p_1}, \ldots, l^{p_n})$-Privacy is proposed. That is, before rating datasets are utilized by the specified recommendation method and data analysts, the user's unique preferences in rating datasets are generalized by less specific data to be indistinguishable. Moreover, the generalized rating dataset can only be utilized through aggregate query frameworks. Although this privacy preservation model can address privacy violation issues in rating datasets, it still has a serious data utility issue that must be improved. For this reason, a new effective privacy preservation model for rating datasets is proposed in this work. Moreover, we propose practical experiments that can measure the effectiveness of the proposed privacy preservation model. The experimental results show that the proposed privacy preservation model is highly effective.

Article information:

Keywords: Multiple Sensitive Attributes, Numerical Quasi-identifiers, Anonymous Data Models, Anonymization Data Models, Data Generalization Models, Rating Datasets and Recommendation Databases

Article history:
Received: September 11, 2022
Revised: October 24, 2022
Accepted: November 6, 2022
Published: November 25, 2022 (Online)

DOI: 10.37936/ecti-cit.2023171.250111

1. INTRODUCTION

Privacy violation issues are serious issues that must be considered when datasets consist of person-specific data, which is private data, to be provided for public use [17] [20]. To address these issues, there have been various privacy preservation models that are proposed in recent decades such as k-Anonymity [15] [22], l-Diversity [10], t-Closeness [7], (k, e)-Anonymous [13] [19] [24], and LKC-Privacy [3] [18]. The common privacy preservation idea of these models is that before datasets are provided for public use, the attributes of datasets are firstly grouped into three groups, i.e., the explicit identifier attributes, the quasi-identifier attributes, and a sensitive attribute. Then, all explicit identifier values are removed. Finally, all unique quasi-identifier values are distorted by using data generalization or data suppression to be indistinguishable. Although these privacy preservation models can address privacy violation issues in datasets, they can only be sufficient to address privacy violation issues in datasets that have a sensitive attribute [11] [12] [14] [16] [20].

Table 1: *An example of rating datasets*

	Artifact attributes		**Personal attributes**		
	Joy Ride	Pachinko	Salary	Age	City
t_1	5	5	$12000	40	NY
t_2	5	5	$15000	48	DC
t_3	2	3	$14000	45	LA
t_4	2	2	$15000	48	DC
t_5	2	2	$16000	45	LA
t_6	4	4	$15000	45	DC
t_7	4	4	$15000	45	DC

[1] The author is with Maejo University, Sansai, Chiangmai, Thailand, 50290, E-mail: surapon_r@mju.ac.th
[1,2,3,4] The authors are with School of Renewable Energy, Maejo University, Sansai, Chiangmai, Thailand, 50290, E-mail: surapon_r@mju.ac.th, kittikorn@mju.ac.th, nigran@mju.ac.th and tanate_c@mju.ac.th
[5] The author is with Maejo University Phrae Campus, Mae Sai, Rong Kwang District, Phrae, Thailand, 54140, E-mail: tammasak@phrae.mju.ac.th

To address this vulnerability of these privacy preservation models, in [4], [8], and [23], the authors propose privacy preservation models that can address privacy violation issues in datasets that have multiple sensitive attributes. Unfortunately, these privacy preservation models are often inappropriate to address privacy violation issues in rating datasets. Because aside from multiple sensitive attributes (the personal attributes), we can see that rating datasets generally have artifact attributes (the quasi-identifier attributes) that only collect numerical data, i.e., the preferred level of interesting artifacts. An example of rating datasets is shown in Table 1. This rating dataset includes seven user profile tuples such that every tuple is constructed from the users' preferred level of their interesting artifacts and their personal information.

Generally, rating datasets are utilized by recommendation methods that are available in real-life systems such as mobile applications, location-based services, e-commerce systems, and websites. Aside from recommendation methods, we see that rating datasets can be further utilized by data analysts. With the data utilization of rating datasets, in [11], [16], and [12], the authors demonstrate that it can lead to privacy violation issues. To address these issues, in [11], the authors recommend using aggregate query frameworks in conjunction with data suppression. They insist all uniquely preferred levels of the user's interesting artifacts are removed before rating datasets will be accessed by aggregate query frameworks. With this privacy preservation idea, in [16], the authors illustrate that although it can address privacy violation issues in rating datasets, it still has data utility issues that must be addressed.

To address this vulnerability of the privacy preservation idea that is proposed in [11], in [16], a privacy preservation framework based on data generalization in conjunction with aggregate query frameworks is presented. That is, the uniquely preferred level of the user's interesting artifacts is generalized by their less specific values to be at least k indistinguishable values and only provided through aggregate query frameworks. For this reason, all possible data utilization conditions of rating datasets through the preferred level of the user's interesting artifacts can guarantee that they always have at least k user profile tuples to be satisfied. However, in [12], the author shows that although the preferred levels of the user's interesting artifacts in rating datasets are generalized to be at least k indistinguishable values, rating datasets still have privacy violation issues that must be addressed. That is, the adversary can use the MAX function in conjunction with the MIN function[1] to disclose the user's personal values that are collected in the personal attributes of rating datasets.

To address privacy violation issues in rating datasets from using the MAX function in conjunction with the MIN function, in [12], $(l^{p_1}, \ldots, l^{p_n})$-Privacy [12] is proposed. With this privacy preservation model, aside from generalizing the uniquely preferred level of the user's interesting artifacts, the number of the different personal values in every personal attribute and the summation of the rating scores of each user profile tuple are also considered in privacy preservation constraints. Although $(l^{p_1}, \ldots, l^{p_n})$-Privacy can address privacy violation issues in rating datasets, it still has data utility loss issues that must be addressed. To rid this vulnerability of $(l^{p_1}, \ldots, l^{p_n})$-Privacy, an effective privacy preservation model for rating datasets is proposed in this work. It will be presented in Section 4.

1.1 Rating datasets

Let $U = \{u_1, u_2, \ldots, u_i\}$ be the set of users. Let $R = RR \cup NULL$, where $RR \subset I^+$, be the set of the possible rating scores. Let $A = \{a_1, a_2, \ldots, a_m\}$ be the set of artifact attributes. Let $P = \{p_1, p_2, \ldots, p_n\}$ be the set of personal attributes. Let $D_{p_z} = \{d_1, d_2, \ldots, d_v\}$ be the data domain of the attribute p_z, where $1 \leq z \leq n$. Let $T = \{t_1, t_2, \ldots, t_i\}$ be the rating dataset such that T is multiset. Every tuple $t_x \in T$, where $1 \leq x \leq i$, represents the profile tuple of the user $u_x \in U$ such that t_x is in the form of $(r_{a_1}, r_{a_2}, \ldots, r_{a_m}, d_{p_1}, d_{p_2}, \ldots, d_{p_n})$, where $r_{a_y} \in R$, $d_{p_z} \in D_{p_z}$, $1 \leq y \leq m$, and $1 \leq z \leq n$. In addition, a higher value of r_{a_y} means a higher preference level for the artifact a_y.

For example, let $U = \{u_1, u_2, u_3, u_4, u_5, u_6, u_7\}$ be the set of users. Let $R = [1, 5] \cup NULL$ be the set of the possible rating scores. Let $A = \{Joy\ Ride, Pachinko\}$ be the set of artifact attributes. Let $P = \{Salary, Age, City\}$ be the set of personal attributes. Moreover, let $D_{Salary} = [10000, 20000]$ be the data domain of the $Salary$ attribute, $D_{Age} = [40, 50]$ be the data domain of the Age attribute, and $D_{City} = \{DC, LA, NY\}$ be the data domain of the $City$ attribute. An example of T is constructed from the given instances to be shown in Table 1. Each tuple t_x of Table 1 represents the profile tuple of the user u_x, where $1 \leq x \leq 7$, e.g., the tuple t_3 is the profile tuple of the user u_3 such that he/she prefers "$Pachinko$" more than "$Joy\ Ride$". Moreover, the user u_3 is 45 years old, \$14000 is his/her salary, and he/she lives in LA.

The organization of this work is as follows. At first, the related works are discussed in Section 2. Subsequently, the motivation of this work is presented in Section 3. Then, the proposed privacy preservation

[1] The MAX and MIN functions return a value as the query result such that: the MAX function returns the query result as the maximum value of a set of specific values. The MIN function returns the query result as the minimum value of a set of particular values. They are determined to support the data domains as numeric, character, unique-identifier, date-time, and etc., This is referenced from the website: https://msdn.microsoft.com/en-us/library/ms187751.aspx.

Fig.1: *The workflow of recommendation methods*

model is presented in Section 4. Then, Section 5. is devoted to discussing the experimental results. Finally, the conclusion and future work are discussed in Section 6.

2. RELATED WORK

2.1 Recommendation methods

Rating datasets are generally proposed for use in recommendation methods. The workflow of using rating datasets in recommendation methods is shown in Figure 1. Recommendation methods use the profile tuple of the target user for considering the appropriately recommended information from rating datasets to the target user. One of the well-known recommendation methods used in several real-life systems is Item-to-Item collaborative filtering [9] [21]. The recommendation idea of this method is that *"The user who likes artifact X, he/she might also like artifact Y because the users who like artifact X also like artifact Y"*. For example, from the data statistic, the users view the cigarette detail, they often view the disposable diaper detail. For this reason, if *Harry* views the cigarette detail, the method recommends the disposable diaper to *Harry*. Although this recommendation idea is more effective in terms of recommending the appropriate artifacts to the target user, it is reported in [2] and [11] that it can lead to privacy violation issues. That is, if the adversary enough has background knowledge about the target user, the adversary can use it to fake the method for snooping or tracking the target user.

2.2 Provided datasets

Aside from recommendation methods, we can see that rating datasets are often provided to data analysts with an appropriate business reason such as constructing business reports for improving business services and management policies. For privacy preservation, before rating datasets are provided to data analysts, all explicit identifier values of users are removed. Unfortunately, in [9], [10], [21], and [22], the authors demonstrate that although provided rating datasets do not contain any explicit identifier values of users, they still have privacy violation issues that must be addressed. That is, the privacy data of users can be violated by using identity and attribute linkage attacks.

2.2.1 Identity linkage attacks

Suppose that Table 1 is the provided rating dataset, and *Alice* is the target user of the adversary. The adversary wants to reveal *Alice*'s salary. Furthermore, we assume the adversary highly believes that *Alice*'s profile tuple is available in Table 1 and further knows that *Alice* highly prefers "*Pachinko*" to "*Joy Ride*". For this situation, the adversary can deduce that \$14000 is *Alice*'s salary because only the tuple t_3 can match the adversary's background knowledge about *Alice*.

To address privacy violation issues from using identity linkage attacks in provided rating datasets, in [22], the author suggests that aside from removing all explicit identifier values, the unique quasi-identifier values, the unique users' preferred level of interesting artifacts, must be distorted by using data generalization or data suppression to ensure that there are at least k indistinguishable tuples. This privacy preservation idea is called a k-anonymization model. The workflow of k-anonymization models is shown in Figure 2. Thus, after rating datasets satisfy anonymization constraints, they can guarantee that all possible re-identifications through the users' preferred level of interesting artifacts always have at least k satisfied tuples. In this situation, privacy violation issues in rating datasets seem to be impossible. Unfortunately, in [10], the authors indeed demonstrate that although rating datasets can guarantee that all possible re-identification conditions have at least k tuples to be satisfied, they still have privacy violation issues that must be addressed to present attribute linkage attacks.

2.2.2 Attribute linkage attacks

Suppose that Table 1 is the provided rating dataset, and *Emma* is the target user of the adversary. The adversary wants to reveal *Emma*'s salary. Furthermore, we assume that the adversary highly believes that *Emma*'s profile tuple is available in Table 1 and he/she further knows that *Emma* is 48 years old. With this adversary's background knowledge, the adversary can see that it matches t_2 and t_4 of Table 1. Moreover, the adversary can observe that both matched tuples only relate to \$15000 as the salary. In this situation, the adversary can infer that \$15000 is *Emma*'s salary. To address privacy violation issues in provided rating datasets from attribute linkage at-

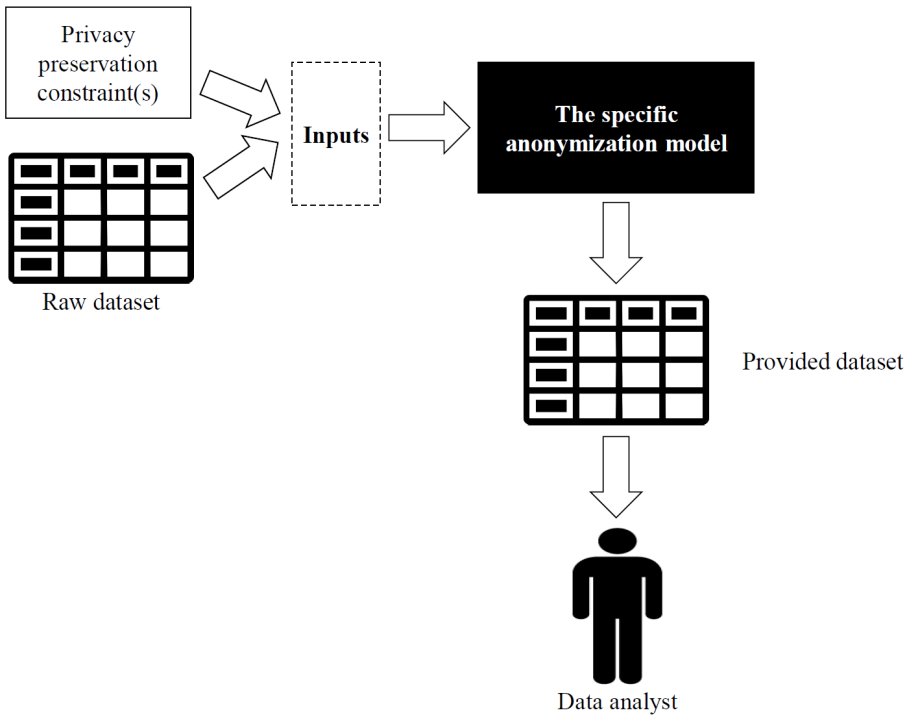

Fig.2: *The workflow of anonymization models*

tacks, in [10], the authors recommend that aside from suppressing or generalizing the unique quasi-identifier values (the users' preferred level of interesting artifacts), the number of the distinct sensitive values in every sensitive attribute must also be considered in privacy preservation constraints. That is, every group of indistinguishable quasi-identifier values must collect at least l different sensitive values. For this reason, after provided datasets satisfy privacy preservation constraints that are proposed in [10], they can guarantee that all possible query conditions through the quasi-identifier values always have at least l differently related sensitive values. Unfortunately, to the best of our knowledge about data generalization and data suppression, they often have data utility issues that must be addressed. For this reason, an alternate privacy preservation model, the aggregate query framework, is proposed. It is presented in the next section.

2.3 Aggregate query frameworks

With privacy preservation that is based on aggregate query frameworks, rating datasets are not directly provided to data analysts. That is, data analysts can only utilize rating datasets through query frameworks that have limitations in that the query result must be queried from a personal attribute by using an aggregate query function and its query condition is determined via the artifact attributes. The workflow of privacy preservation for rating datasets based on aggregate query frameworks is shown in Figure 3.

For example, Table 1 is the provided rating dataset. An example query is shown in *Query 1*.

- *Query 1: "SELECT AVERAGE(Age) AS Answer FROM Table 1 WHERE Pachinko* $\geq$ *3"*

The query result of *Query 1* is 44.33 as the average age. It is the aggregate query answer. Thus, it does not seem to have any privacy violation issues. Unfortunately, in [11], [12], and [16], the authors demonstrate that aggregate query frameworks still have privacy violation issues that must be addressed. That is, the privacy data of users in provided rating datasets could be violated by using counting attacks.

2.3.1 Counting attacks

Let Table 1 be the provided rating dataset. Let *Alice* be the target user of the adversary. We assume the adversary knows that *Alice* is 45 years old and lives in *LA*. Moreover, the adversary highly believes that *Alice*'s tuple profile is available in Table 1, and the adversary wants to reveal *Alice*'s salary. For revealing *Alice*'s salary, the adversary first uses the *COUNT* function to determine the risk query conditions. That is, the query conditions of Table 1 only have a user profile tuple that can be satisfied. We suppose that in a trial-and-error-process using the *COUNT* function, the adversary can eventually determine the query condition "*Pachinko* = 3" is the risk query condition because this query condition only has the user profile tuple t_3 that can be satisfied, as shown in *Query 2*.

- *Query 2 : "SECLECT COUNT(*) AS Rows FROM Table 1 WHERE Pachinko = 3"*

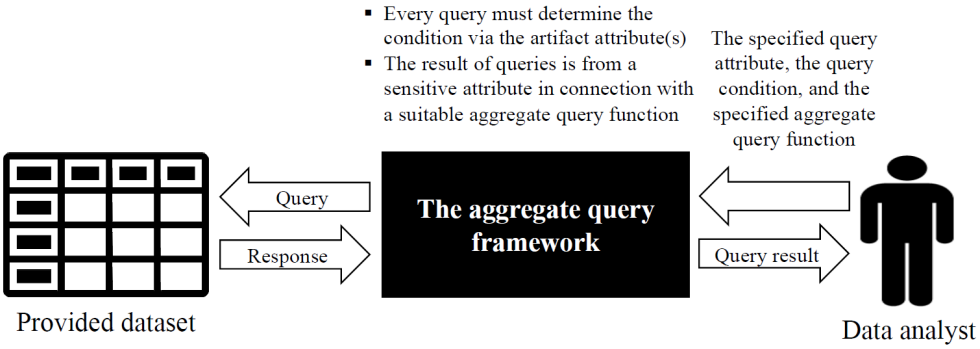

Fig.3: *The workflow of aggregate query frameworks*

Subsequently, the adversary uses the determined risk query condition in conjunction with the MAX function and his/her background knowledge for verifying *Alice*'s age and *Alice*'s city, as shown in *Queries 3* and *4* respectively.

- *Query 3 : "SECLECT MAX (Age) AS Answer FROM Table 1 WHERE Pachinko = 3"*
- *Query 4 : "SECLECT MAX (City) AS Answer FROM Table 1 WHERE Pachinko = 3"*

The query result of *Query 3* is 45 as the maximum age, and LA is the query result that is returned from *Query 4*. With these query results, the adversary can be highly confident that they are not the aggregate query answer. Moreover, the adversary can see that these query results match his/her background knowledge about *Alice*. For this situation, the adversary can be highly confident that the user profile tuple is satisfying the query condition "$Pachinko = 3$" to be *Alice*'s profile tuple. Finally, the adversary can use this query condition in conjunction with the MAX function to reveal *Alice*'s salary, $14000, as shown in *Query 5*.

- *Query 5 : "SECLECT MAX (Salary) AS Answer FROM Table 1 WHERE Pachinko = 3"*

In this situation, we can conclude that although provided rating datasets do not include any explicit identifier value of users and cannot be directly permitted for data utilization, they still have privacy violation issues that must be addressed. To address these privacy violation issues in provided rating datasets, a privacy preservation model based on data anonymization in conjunction with aggregate query frameworks to be proposed, it will be presented in Section 2.4

2.4 Aggregate query frameworks based on data anonymization

To address privacy violation issues in provided rating datasets from the adversary employing counting attacks, in [11] and [16], the authors propose aggregate query frameworks that are based on data anonymization. The workflow of privacy preservation is proposed in [11] and [16], it is shown in Figure 4. That is, aside from rating datasets that are provided by aggregate query frameworks, the unique users' preferred levels of interesting artifacts are further distorted by using data generalization or data suppression to guarantee the existence of at least k indistinguishable tuples. However, in [12], the author demonstrates that although the unique users' preferred levels of interesting artifacts are generalized or suppressed to be at least k indistinguishable tuples, provided rating datasets still have privacy violation issues that must be addressed because the privacy data of users can still be revealed by using identical attacks.

2.4.1 Identical attacks

Let Table 2 be the provided rating dataset. Given *Bob* is the target user. The adversary wants to reveal *Bob*'s salary from Table 2. We assume the adversary knows that *Bob* is 45 years old and lives in DC. Moreover, the adversary strongly believes that *Bob*'s profile tuple is available in Table 2. For revealing *Bob*'s salary, the adversary first uses the MAX and MIN functions in conjunction with his/her background knowledge about *Bob* to determine the desired risk query condition. That is, an arbitrary query condition of Table 2 has the query results from using the MAX and MIN functions that match the adversary's background knowledge. We suppose that after the trial-and-error-process with the MAX and MIN functions, the adversary can eventually determine the query condition "$Joy\ Ride = 4$" to be the risk query condition because the query results from using the MAX and MIN functions of this query condition are identical and match the adversary's background knowledge. For this situation, the adversary can be highly confident that the user profile tuple is satisfying the query condition "$Joy\ Ride = 4$" to be *Bob*'s profile tuple. Thus, the adversary can use this query condition in conjunction with the MAX and MIN functions to reveal *Bob*'s salary, $15000.

In this situation, we can conclude that although the unique users' preferred levels of interesting artifacts in proposed rating datasets are generalized or suppressed to have at least k indistinguishable tuples and they cannot be used directly, they still have privacy violation issues that must be addressed. That is,

Table 2: *An example of data anonymization versions of Table 1*

	Artifact attributes		**Personal attributes**		
	Joy Ride	Pachinko	Salary	Age	City
t_1	5	5	\$12000	40	NY
t_2	5	5	\$15000	48	DC
t_3	2	{2,3}	\$14000	45	LA
t_4	2	{2,3}	\$15000	48	DC
t_5	2	{2,3}	\$16000	45	LA
t_6	4	4	\$15000	45	DC
t_7	4	4	\$15000	45	DC

the privacy data of users can still be violated by using identical attacks. To address these privacy violation issues, in [12], the author proposes $(l^{p_1}, \ldots, l^{p_n})$-Privacy. With this privacy preservation model, aside from distorting the unique users' preferred levels of interesting artifacts, the distortedly preferred levels of interesting artifacts must relate to the person-specific values which are available in every personal attribute p_z, where $1 \leq z \leq n$, to have at least l^{p_z} different values. For this reason, after rating datasets satisfy $(l^{p_1}, \ldots, l^{p_n})$-Privacy constraints, they can guarantee that all possible query conditions always have at least l^{p_z} different values in every personal attribute p_z. More details of $(l^{p_1}, \ldots, l^{p_n})$-Privacy will be explained in Section 3.

3. MOTIVATION

In [12], $(l^{p_1}, \ldots, l^{p_n})$-Privacy is proposed. With this privacy preservation model, rating datasets can satisfy privacy preservation constraints with the following steps:

- In the first step, the rating scores of each tuple in T are summed.
- In the second step, the tuples of T are re-sorted in ascending order by their summed values.
- In the third step, the tuples of T are partitioned by considering their order such that every personal attribute p_z, where $1 \leq z \leq n$, of each partition of T must include at least l^z difference values.
- In the fourth step, the unique rating scores of every partition are generalized by their less specific values to be indistinguishable.
- Finally, the generalized data version of T is provided through an aggregate query framework.

An example of privacy preservation in rating datasets based on $(l^{p_1}, \ldots, l^{p_n})$-Privacy constraints. Let Table 1 be the specified original rating dataset such that the data must be accessed through an aggregate query framework. Let $l^{Salary} = 2$, $l^{Age} = 2$, and $l^{City} = 2$, be the specified privacy preservation constraint for the attributes *Salary*, *Age*, and *City* respectively. With these specified privacy preservation constraints, a provided data version of Table 1 is shown in Table 3. Aside from Table 3, we can see that Table 4 is also a provided rating dataset version of Table 1 such that it is satisfied by the specified privacy preservation constraints. For this reason, we can conclude that a particular rating dataset and a specified $(l^{p_1}, \ldots, l^{p_n})$-Privacy constraint generally have multiple possibly provided data versions. Moreover, with Tables 3 and 4, we can see that their unique users' preferred levels of interesting artifacts are generalized by the differently less specified values. For this reason, they could be different data utilizations. Thus, only the provided data version is highly data utilities to be desired. However, to the best of our knowledge about $(l^{p_1}, \ldots, l^{p_n})$-Privacy, it cannot guarantee that the returnedly generalized data version, the provided data version, is the data version that has the highest data utilities. To rid this vulnerability of $(l^{p_1}, \ldots, l^{p_n})$-Privacy, a newly effective privacy preservation model for rating datasets based on $(l^{p_1}, \ldots, l^{p_n})$-Privacy constraints to be proposed in this work. It is presented in the next section.

4. EFFECTIVE $(L^{P_1}, \ldots, L^{P_N})$-PRIVACY

Although $(l^{p_1}, \ldots, l^{p_n})$-Privacy can address privacy violation issues in proposed ratting datasets, it still has data utility issues that must be addressed. To do that, an enhanced privacy preservation model of $(l^{p_1}, \ldots, l^{p_n})$-Privacy is proposed in this section. With the proposed model, aside from privacy preservation constraints, the data utility is also maintained as much as possible.

4.1 Basic definitions

Definition 1 (Equivalence classes) Let l^{p_z} be the privacy preservation constraint for the personal attribute $p_z \in P$ such that it is a positive integer that is equal to or greater than 2, i.e., $l^{p_z} \in I^+$ and $l^{p_z} \geq 2$. Let ec be an arbitrary equivalence class of T such that the unique users' preferred levels of interesting artifacts in ec are generalized by their range to be indistinguishable. Moreover, every p_z of ec must collect at least l^{p_z} different personal values.

Definition 2 (The error of equivalence classes) Let $f_{\mathbb{U}}(ec_g[a_y])$ and $f_{\mathbb{L}}(ec_g[a_y])$, where $1 \leq y \leq m$, be the upper and lower bound of a_y of equivalence class ec_g. Therefore, the error of ec_g can be defined by $f_{\mathbb{E}}(ec_g) = \sum_{y=1}^{m} f_{\mathbb{U}}(ec_g[a_y]) - f_{\mathbb{L}}(ec_g[a_y])$. Higher values of $f_{\mathbb{E}}(ec_g)$ mean that ec_g has less data utility.

For example, let $\{2, 4, 5\}$ be the set of the rating scores that are available in the artifact attribute a_{y_1} of the equivalence class ec_g. Moreover, let $\{2, 3, 4\}$ be the set of the rating scores that are available in the artifact attribute a_{y_2} of the equivalence class ec_g. For this situation, the generalized value of a_{y_1} is $2-5$, and $2-4$ is the generalized value of a_{y_2}. The upper bound, $f_{\mathbb{U}}(ec_g[a_{y_1}])$, of the artifact attribute a_{y_1} of the equivalence class ec_g is 5. The lower bound, $f_{\mathbb{L}}(ec_g[a_{y_1}])$, of the artifact attribute a_{y_1} of the equivalence class ec_g is 2. The upper bound, $f_{\mathbb{U}}(ec_g[a_{y_2}])$, of the artifact attribute a_{y_2} of the equivalence class ec_g is 4. The

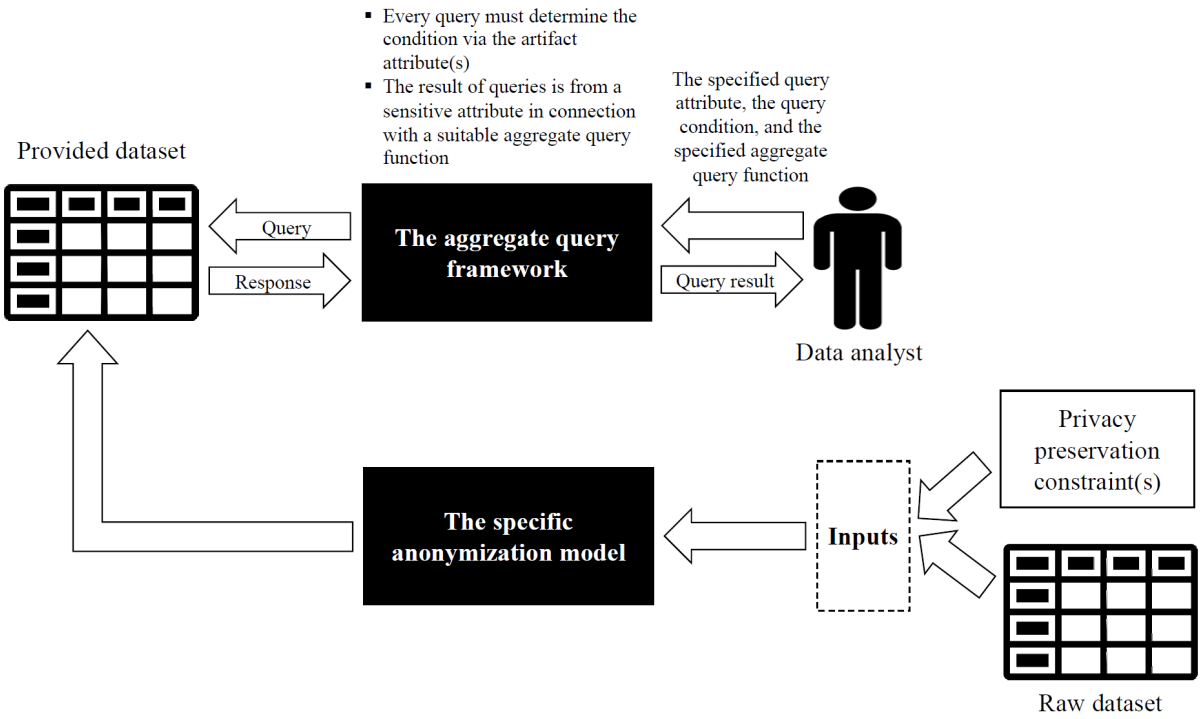

Fig.4: *The workflow of privacy preservation based on data anonymization in conjunction with aggregate query frameworks*

lower bound, $f_{\mathbb{L}}(ec_g[a_{y_2}])$, of the artifact attribute a_{y_2} of the equivalence class ec_g is 2. Therefore, the error of ec_g is 5, i.e., $f_{\mathbb{E}}(ec_g) = (5-2)+(4-2)$.

Definition 3 ($(l^{p_1},\ldots,l^{p_n})$-Privacy) Let T' be the provided data version of T such that T' is constructed by the set of equivalence classes as ec_1, ec_2, …, ec_e. Without loss of generality, $ec_1 \cup ec_2 \cup \cdots \cup ec_e = T$ and $ec_1 \cap ec_2 \cap \cdots \cap ec_e = \emptyset$. Furthermore, the summation of all tuples of ec_g must be less than or equal to the summation of all tuples of the partition ec_{g+1}, where $1 \leq g \leq e$.

Definition 4 (Error of datasets) The penalty cost of T' can be defined from the error of its equivalence classes such that $f_{\mathbb{D}}(T') = \sum_{g=1}^{e} f_{\mathbb{E}}(ec_g)$. Higher values of $f_{\mathbb{D}}(T')$ mean that T' is less data utility.

Definition 5 (Effective $(l^{p_1},\ldots,l^{p_n})$-Privacy) Let T'_{Eff} be the effectively provided data version of T such that T'_{Eff} is satisfied by the following limitations:

- T'_{Eff} is satisfied by Definition 3,
- $f_{\mathbb{D}}(T'_{Eff})$ is minimized.
- The number of equivalence classes of T'_{Eff} is maximized.

For example, with Table 3, the error of $f_{\mathbb{E}}(ec_1)$ is 0. The error of $f_{\mathbb{E}}(ec_2)$ is 5. Therefore, the error of this table, $f_{\mathbb{D}}$(Table 3), is 5. With Table 4, the error of $f_{\mathbb{E}}(ec_1)$ is 1. The error of $f_{\mathbb{E}}(ec_2)$ is 2, so, $f_{\mathbb{D}}$(Table 4) is 3. Therefore, Table 4 is the T'_{Eff} version of Table 1 such that it satisfies $l^{Salary} = 2$, $l^{Age} = 2$, and $l^{City} = 2$.

The T'_{Eff} version of T can be constructed by the following step:

- At first, the set of all possible equivalence classes, denoted as EC, of T is constructed such that every $ec_g \in EC$ is satisfied by Definition 1.
- Then, the set of all possibly generalized data versions, denoted as T^*, of T is constructed from EC, i.e., every $T' \in T^*$ is satisfied by Definition 3.
- Then, T'_{Eff} of T is determined such that $f_{\mathbb{D}}(T'_{Eff}) \leq f_{\mathbb{D}}(T'_1) \leq \ldots \leq f_{\mathbb{D}}(T'_s)$, where $T'_{Eff} \cup T'_1 \cup \cdots \cup T'_{1s} = T^*$. Moreover, the number of equivalence classes of T'_{Eff} is maximized.
- Finally, T'_{Eff} is provided by an aggregate query framework.

Why must we consider the number of equivalence classes in effective $(l^{p_1},\ldots,l^{p_n})$-Privacy constraints? In some situations, the particular rating dataset based on a specified $(l^{p_1},\ldots,l^{p_n})$-Privacy constraint to have multiple provided data versions that have the same error but they have a different number of equivalence classes. Therefore, only the provided data version of the particular rating dataset has the maximized number of equivalence classes to be desired because the size of its equivalence classes is small.

4.2 Data utility metrics

After rating datasets are satisfied by privacy preservation constraints, they often lead to being considered as having high security in terms of privacy preservation. However, they generally lose some data utility. For this reason, the data utility metric is necessary. Since privacy preservation models are

Table 3: *A provided data version of Table 1 satisfies $(l^{p_1}, \ldots, l^{p_n})$-Privacy constraints, where $l^{Salary} = 2$, $l^{Age} = 2$, and $l^{City} = 2$*

	Artifact attributes		**Personal attributes**			**Equivalence class**
	Joy Ride	Pachinko	Salary	Age	City	
t_4	2	2	$15000	48	DC	1
t_5	2	2	$16000	45	LA	
t_3	2-5	3-5	$14000	45	LA	2
t_6	2-5	3-5	$15000	45	DC	
t_7	2-5	3-5	$15000	45	DC	
t_1	2-5	3-5	$12000	40	NY	
t_2	2-5	3-5	$15000	48	DC	

Table 4: *A provided data version of Table 1 satisfies effective $(l^{p_1}, \ldots, l^{p_n})$-Privacy constraints, where $l^{Salary} = 2$, $l^{Age} = 2$, and $l^{City} = 2$*

	Artifact attributes		**Personal attributes**			**Equivalence class**
	Joy Ride	Pachinko	Salary	Age	City	
t_4	2	2-3	$15000	48	DC	1
t_5	2	2-3	$16000	45	LA	
t_3	2	2-3	$14000	45	LA	
t_6	4-5	4-5	$15000	45	DC	2
t_7	4-5	4-5	$15000	45	DC	
t_1	4-5	4-5	$12000	40	NY	
t_2	4-5	4-5	$15000	48	DC	

proposed, there are several well-known data utility metrics to be proposed such as Normalized certainty penalty for numeric data domain (NCP) [5], Discernibility metric DM) [1], and Relative error [24].

4.2.1 Normalized certainty penalty for numeric data domain (NCP)[5]

The NCP metric is one of the well-known data utility metrics that can be used to define the penalty cost or the data utility of $ec_g \in EC$. Higher values of NCP of ec_g mean that ec_g is more generalized or has less data utility. The NCP cost of ec_g can be defined by Equation 1.

$$f_{\mathbb{NCP}}(T', ec_g) = \sum_{y=1}^{m} \frac{f_{\mathbb{U}}(ec_g[a_y]) - f_{\mathbb{L}}(ec_g[a_y])}{f_{\mathbb{U}}(T'[a_y]) - f_{\mathbb{L}}(T'[a_y])} \quad (1)$$

The penalty cost of T' can be defined by Global Certainty Penalty (GCP), as shown in Equation 2. Also, higher values of GCP mean that T' is more generalized or has less data utility.

$$f_{\mathbb{GCP}}(T') = \frac{1}{m \cdot |T'|} \cdot \sum_{g=1}^{e} |ec_g| \cdot f_{\mathbb{NCP}}(T', ec_g) \quad (2)$$

4.2.2 Discernibility metric (DM) [1]

The DM metric is a data utility metric that can be used to define the penalty cost or the data utility of T'. With the DM metric, the penalty cost of T' is dependent on the size of its equivalence classes. The DM cost of T' can be defined by Equation 3. Higher values of DM mean that T' is less data utility.

$$f_{\mathbb{DM}}(T') = \sum_{g=1}^{e} |ec_g|^2 \quad (3)$$

4.2.3 Relative error [24]

The relative error is another metric that can be used to define the penalty cost of T'. The relative error of T' depends on the different values of T' and its original dataset T. The relative error metric is shown in Equation 4.

$$f_{\mathbb{RE}}(\varphi, \varphi_0) = \frac{|\varphi - \varphi_0|}{\varphi} \quad (4)$$

Where,

- φ is the original result such that it is available in T.
- φ_0 is the relatively experimental result such that it relates to φ and it is available in T'.

5. EXPERIMENT

In this section, the effectiveness of the proposed privacy preservation model is evaluated by comparison with the comparable privacy preservation model that is proposed in [12].

5.1 Experimental setup

All experiments are conducted on an Intel(R) Xeon(R) Silver 4110 @2.10 GHz CPU with 16 GB memory and 1 TB HDD running Microsoft Windows

10 64-bit professional edition. All implementations are built and executed on Microsoft Visual Studio 2017 Community Edition in conjunction with Microsoft SQL Server 2019.

Moreover, every experiment is evaluated on the "$ml-10m$" dataset which is proposed by the "$GroupLens$" recommendation system [6]. It is described by the rating scores in the range between 1 and 5. Furthermore, it contains 10681 movies and 71567 user profiles. Every user profile is represented by "$User\ ID, Age, Gender, Occupation$, and $Zipcode$". In addition, each movie is rated by at least one user, and each user rates at least 20 movies.

To conduct effective experiments, only the top ten movies that have the highest number of ratings and the users who rated those movies are selected, though the trends in the effectiveness can still be seen. Thus, the dataset only retains 4230 ratings from 423 users for 10 movies. Furthermore, only the personal attributes "$Age, Occupation$, and $Zipcode$" are available in the experimental dataset.

5.2 Experimental results and discussions

5.2.1 The effectiveness based on data generalization

In this section, the effect of data generalization in provided rating datasets is evaluated by using the GCP metric.

In the first experiment, the effect of GCP penalty costs based on the number of personal attributes is evaluated. For experiments, the number of artifact attributes is fixed at 10. The privacy preservation constraints for every personal attribute are set at 2. The number of personal attributes varies from 1 to 3. From the experimental results shown in Figure 5, we can conclude that the number of personal attributes has an effect on GCP penalty costs for all experimentally provided rating datasets. When the number of personal attributes is increased, the GCP penalty cost of all experimentally provided rating datasets is also increased. The cause of increasing GCP penalty costs of all experimentally provided rating datasets is that the size of their equivalence classes and the range of generalized rating scores increases when the number of personal attributes is increased. Moreover, we can observe that the number of personal attributes influences the GCP penalty cost of the experimentally provided rating datasets which are constructed by the comparative privacy preservation model higher than the proposed privacy preservation model.

In the second experiment, the effect of GCP penalty costs based on the number of artifact attributes is evaluated. For experiments, the number of personal attributes is fixed at 3. The privacy preservation constraints for every personal attribute are set at 2. The number of artifact attributes varies from 2 to 10. From the experimental results shown in Figure 6, we can conclude that the number of artifact attributes has an effect on GCP penalty costs for all experimentally provided rating datasets. When the number of artifact attributes is increased, the GCP penalty cost of all experimentally provided rating datasets also increases. The cause of increasing GCP penalty costs is that the number of generalized rating scores available in the experimentally provided rating datasets increases when the number of personal attributes is increased. Moreover, we can observe that the number of artifact attributes influences the GCP penalty cost of the experimentally provided rating datasets which are constructed by the comparative privacy preservation model higher than the proposed privacy preservation model.

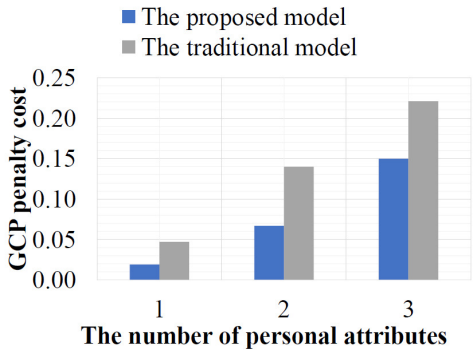

Fig.5: *The GCP penalty costs based on the number of personal attributes.*

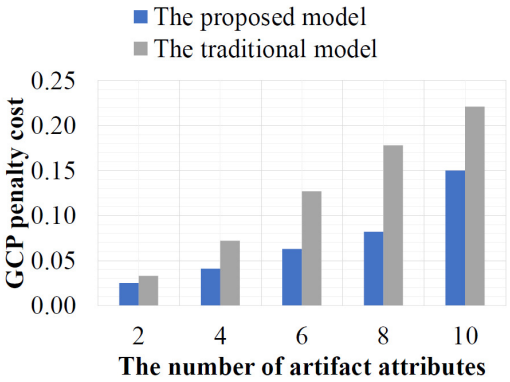

Fig.6: *The GCP penalty costs based on the number of artifact attributes.*

In the final experiment, the effect of GCP penalty costs based on the privacy preservation constraints $l^{p_1}, \ldots, l^{p_n}$ is evaluated. For experiments, the number of personal attributes is fixed at 3. The number of artifact attributes is fixed at 10. The privacy preservation constraints $l^{p_1}, \ldots, l^{p_n}$ varies from 1 to 5. From the experimental results shown in Figure 7, we can conclude that the privacy preservation constraints $l^{p_1}, \ldots, l^{p_n}$ also have an effect of GCP penalty costs to all experimentally provided rating datasets. When the privacy preservation constraints $l^{p_1}, \ldots, l^{p_n}$ are increased, the GCP penalty cost for all experimentally provided rating datasets also increases. The cause for increasing GCP penalty costs

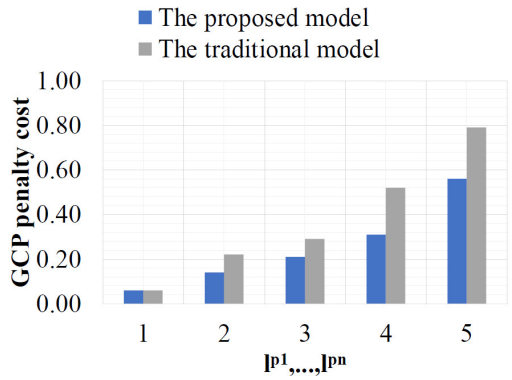

Fig.7: *The GCP penalty costs based on* $l^{p_1}, \ldots, l^{p_n}$.

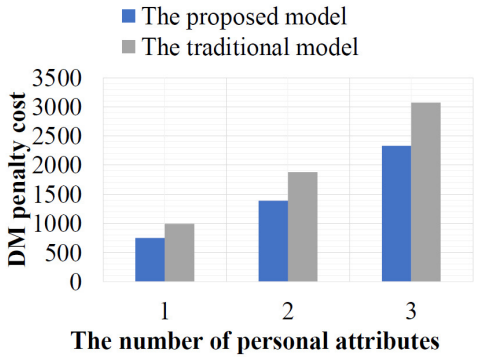

Fig.8: *The DM penalty costs based on the number of personal attributes.*

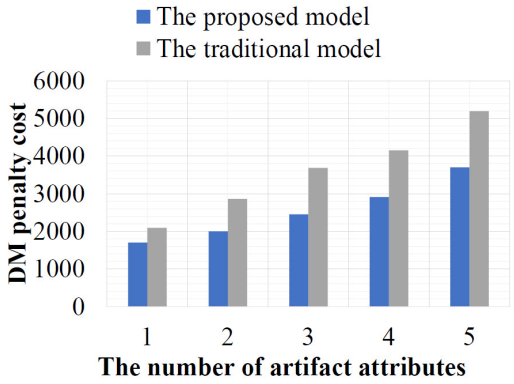

Fig.9: *The DM penalty costs based on the number of artifact attributes.*

is that the number of generalized rating scores and the range of generalized rating scores available in the experimentally provided rating datasets increases when the value of privacy preservation constraints $l^{p_1}, \ldots, l^{p_n}$ are increased. In addition, the experimental results show an effect of GCP penalty costs when the value of $l^{p_1}, \ldots, l^{p_n}$ is 1 because the rating scores are available in the experimental dataset to be the same accidentally. Moreover, we can observe that the privacy preservation constraints $l^{p_1}, \ldots, l^{p_n}$ also influence the GCP penalty cost of the experimentally provided rating datasets which are constructed by the comparative privacy preservation model more than the proposed privacy preservation model.

5.2.2 The effectiveness based on the size of equivalence classes

In this section, the effect of the size of equivalence classes in provided rating datasets is evaluated by using the DM metric.

In the first experiment, the effect of DM penalty costs based on the number of personal attributes is evaluated. For experiments, the number of artifact attributes is fixed at 10. The privacy preservation constraints for every personal attribute are set at 2. The number of personal attributes varies from 1 to 3. From the experimental results shown in Figure 8, we conclude that the number of personal attributes has an effect on DM penalty costs to all experimentally provided rating datasets when the number of personal attributes is increased, the DM penalty cost for all experimentally provided rating datasets also increases. Moreover, we can observe that the number of personal attributes influences the DM penalty cost of the experimentally provided rating datasets which are constructed by the comparative privacy preservation model more than the proposed privacy preservation model.

In the second experiment, the effect of DM penalty costs based on the number of artifact attributes is evaluated. For experiments, the number of personal attributes is fixed at 3. The privacy preservation constraints for every personal attribute are set at 2. The number of artifact attributes varies from 2 to 10. From the experimental results shown in Figure 9, we conclude that the number of artifact attributes has an effect on DM penalty costs for all experimentally provided rating datasets. When the number of artifact attributes is increased, the DM penalty cost of all experimentally provided rating datasets also increases. The cause of increasing DM penalty costs is that the number of generalized rating scores available in the experimentally provided rating datasets is increased when the number of personal attributes is increased. In this situation, the size of their equivalence classes in the experimentally provided rating datasets also increases. Moreover, we can observe that the number of artifact attributes influences the DM penalty cost of the experimentally provided rating datasets which are constructed by the comparative privacy preservation model more than the proposed privacy preservation model.

In the final experiment, the effect of DM penalty costs based on the privacy preservation constraints $l^{p_1}, \ldots, l^{p_n}$ is evaluated. For experiments, the number of personal attributes is fixed at 3. The number of artifact attributes is fixed at 10. The privacy preservation constraints $l^{p_1}, \ldots, l^{p_n}$ are varied from 1 to 5. From the experimental results shown in Figure 10, we conclude that the privacy preserva-

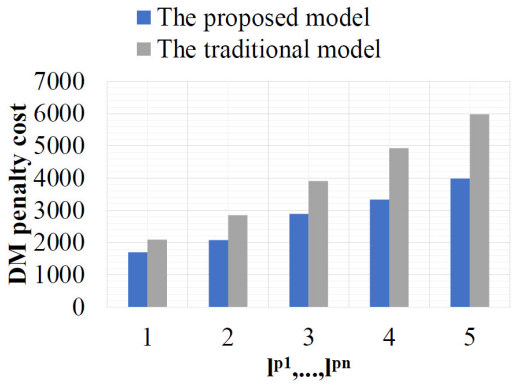

Fig.10: *The DM penalty costs based on* $l^{p_1}, \ldots, l^{p_n}$.

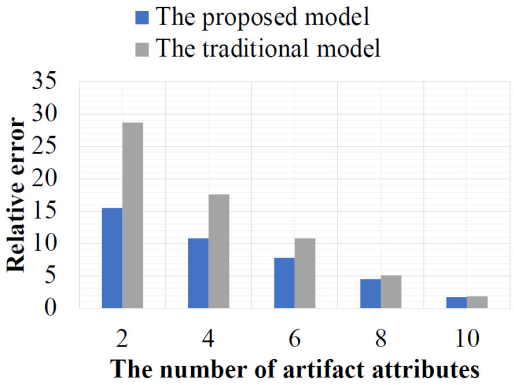

Fig.11: *The relative error based on the OR query operation*

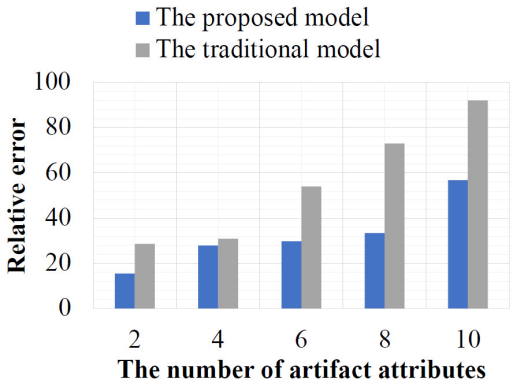

Fig.12: *The relative error based on the AND query operation*

tion constraints $l^{p_1}, \ldots, l^{p_n}$ also have an effect of DM penalty costs for all experimentally provided rating datasets. When the privacy preservation constraints $l^{p_1}, \ldots, l^{p_n}$ are increased, the DM penalty cost for all experimentally provided rating datasets also increases. The cause of increasing DM penalty costs is that the size of their equivalence classes available in the experimentally provided rating datasets is increased when the value of privacy preservation constraints $l^{p_1}, \ldots, l^{p_n}$ are increased. In addition, the experimental results show an effect of DM penalty costs when the value of $l^{p_1}, \ldots, l^{p_n}$ is 1 because the rating scores are available in the experimental dataset to be the same accidentally. Moreover, we can observe that the privacy preservation constraints $l^{p_1}, \ldots, l^{p_n}$ also influence the DM penalty cost of the experimentally provided rating datasets which are constructed by the comparative privacy preservation model more than the proposed privacy preservation model.

5.2.3 The effectiveness based on query results

In this section, the effect of query results based on the OR operation, the AND operation, and the range of queries are evaluated by using the relative error metric. From the experimental results shown in Figure 11, 12, and 13, we observe that the trend of query results that are queried by the OR operation and the range of queries is different from the AND operation. With the OR operation and the range of queries, when the number of artifact attributes is increased, the data utility of query results is also increased, but the query results of the AND operation contrast from the OR operation and the range of queries. When the number of artifact attributes is increased, the query results between the original dataset and its experimentally provided data versions are queried by the AND operation to be more different. Furthermore, we can observe that all experimental results of the proposed model have a penalty cost of the relative error less than the comparative model.

From the experimental results that are proposed in Sections 5.2.1, 5.2.2, and 5.2.3, we conclude that the proposed privacy preservation model has effectiveness in terms of data generalization, constructing equivalence classes, and query results more than the comparative privacy preservation model.

6. CONCLUSION AND FUTURE WORK

In this work, privacy violation issues in rating datasets are identified. To address these issues, we propose a new privacy preservation model which is based on (l^{p_1}, ..., l^{p_n})-Privacy constraints. Moreover, we show the experiments whose results indicate the proposed model to be an effective privacy preservation model for rating datasets.

In addition, although the proposed model can address privacy violation issues in rating datasets, the adversary may identify a vulnerability of the proposed model in the future. The adversary could then use the identified vulnerability to violate the privacy data of users in rating datasets. Therefore, suitable privacy preservation models that can address the newly identified privacy violation issues of the adversary must also be proposed in the future.

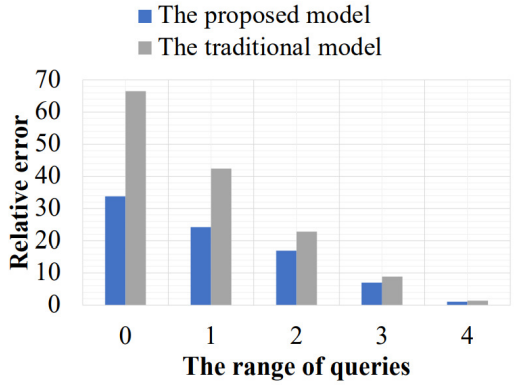

Fig.13: *The relative error based on the range of queries*

7. COMPLIANCE WITH ETHICAL STANDARDS

Conflict of interest Authors declare that they have no conflict of interest. **Ethical approval** This paper does not contain any studies with human participants or animals performed by the authors.

References

[1] R.J. Bayardo and R. Agrawal, "Data privacy through optimal k-anonymization," in *21st International Conference on Data Engineering (ICDE'05)*, pp. 217–228, 2005.

[2] J. A. Calandrino, A. Kilzer, A. Narayanan, E. W. Felten and V. Shmatikov, " "You Might Also Like:" Privacy Risks of Collaborative Filtering," *2011 IEEE Symposium on Security and Privacy*, pp. 231-246, 2011.

[3] B.C.M. Fung, M. Cao, B.C. Desai and H. Xu, "Privacy protection for RFID data," in *Proceedings of the 2009 ACM Symposium on Applied Computing, SAC '09*, pp. 1528–1535, 2009.

[4] T. S. Gal, Z. Chen and A. Gangopadhyay, "A privacy protection model for patient data with multiple sensitive attributes," *International Journal of Information Security and Privacy (IJISP)*, vol. 2, no. 3, pp 28–44, 2008.

[5] G. Ghinita, P. Karras, P .Kalnis and N. Mamoulis, "A framework for efficient data anonymization under privacy and accuracy constraints," *ACM Transactions on Database Systems*, vol. 34, Issue 2, no. 9, pp. 1-47, 2009.

[6] F.M. Harper and J.A. Konstan, " The Movielens Datasets: History and Context," *ACM Transactions on Interactive Intelligent Systems*, vol. 5, Issue 4, no. 19, pp. 1–19, 2015.

[7] P. Tirelli, N. A. Borghese, F. Pedersini, G. Galassi and R. Oberti, "Automatic monitoring of pest insects traps by Zigbee-based wireless networking of image sensors," *2011 IEEE International Instrumentation and Measurement Technology Conference*, pp. 1-5, 2011.

[8] Z. Li and X. Ye, "Privacy protection on multiple sensitive attributes," *International Conference on Information and Communications Security*, pp. 141–152, 2007.

[9] G. Linden, B. Smith and J. York, "Amazon.com recommendations: item-to-item collaborative filtering," in *IEEE Internet Computing*, vol. 7, no. 1, pp. 76-80, Jan.-Feb. 2003

[10] A. Machanavajjhala, D. Kifer, J. Gehrke and M. Venkitasubramaniam, "L-diversity: Privacy beyond k-anonymity," *ACM Transactions on Knowledge Discovery from Data*, vol. 1, no. 1, 2007.

[11] N. Ramakrishnan, B. J. Keller, B. J. Mirza, A. Y. Grama and G. Karypis, "Privacy risks in recommender systems," in *IEEE Internet Computing*, vol. 5, no. 6, pp. 54-63, Nov.-Dec. 2001.

[12] S. Riyana, "(l^{p_1} , ..., l^{p_n})-Privacy: privacy preservation models for numerical quasi-identifiers and multiple sensitive attributes," *Journal of Ambient Intelligence and Humanized Computing*, vol.12, pp. 9713–9729, 2021.

[13] S. Riyana, N. Harnsamut, T. Soontornphand and J. Natwichai, "(k, e)-anonymous for ordinal data," in *2015 18th International Conference on Network-Based Information Systems*, pp. 489–493, 2015.

[14] S. Riyana, N. Ito, T. Chaiya, U. Sriwichai, N. Dussadee, T. Chaichana, R. Assawarachan, T. Maneechukate, S. Tantikul and N. Riyana, "Privacy threats and privacy preservation techniques for farmer data collections based on data shuffling," *ECTI Transactions on Computer and Information Technology (ECTI-CIT)*, vol. 16, no.3, pp. 289–301, 2022.

[15] S. Riyana, S. Nanthachumphu and N. Riyana, "Achieving privacy preservation constraints in missing-value datasets," *SN Computer Science*, vol. 1, no. 227, pp. 1–10, 2020.

[16] S. Riyana and J. Natwichai, "Privacy preservation for recommendation databases," *Service Oriented Computing and Applications*, vol. 12, no. 3-4, pp. 259-273, 2018.

[17] S. Riyana and N. Riyana, "A privacy preservation model for RFID data-collections is highly secure and more efficient than LKC-privacy," in *The 12th International Conference on Advances in Information Technology*, no.15, pp. 1–11, 2021.

[18] S. Riyana and N. Riyana, "Achieving anonymization constraints in high-dimensional data publishing based on local and global data suppressions," *SN Computer Science*, vol. 3, no.1, pp. 1–12, 2022.

[19] S. Riyana, N. Riyana and S. Nanthachumphu, "Enhanced (k,e)-anonymous for categorical data," in *Proceedings of the 6th International Conference on Software and Computer Applica-*

tions, IC- SCA '17, pp. 62–67, 2017.

[20] S. Riyana, N. Riyana and W. Sujinda, "An anatomization model for farmer data collections," *SN Computer Science*, vol. 2, no. 353, pp. 1–11, 2021.

[21] B. Sarwar, G. Karypis, J. Konstan, J. Riedl, "Item-based collaborative filtering recommendation algorithms," in *Proceedings of the 10th International Conference on World Wide Web, WWW '01*, pp. 285–295, 2001.

[22] L. Sweeney, "K-anonymity: A model for protecting privacy," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 10, no. 5, pp. 557–570, 2002.

[23] Y. Ye, Y. Liu, C. Wang, D. Lv and J. Feng, "Decomposition: Privacy preservation for multiple sensitive attributes," in *International Conference on Database Systems for Advanced Applications*, pp. 486–490, 2009.

[24] Q. Zhang, N. Koudas, D. Srivastava and T. Yu, "Aggregate Query Answering on Anonymized Tables," *2007 IEEE 23rd International Conference on Data Engineering*, pp. 116-125, 2007

Kittikorn Sasujit received a B.Eng (Environmental Engineering) in 2004 from Rajamangala University of Technology Lanna, Thailand, and an M. Eng and Ph.D. (Energy Engineering) in 2008 and 2020, respectively, from Chiang Mai University, Thailand. His studies will include biomass technology, wind energy technology, NTP applications for biomass tar removal, and renewable energy.

Nigran Homdoung received a B.S. degree in mechanical engineering from King Mongkut 's University of Technology Thonburi (KMUTT), Thailand, in 2001. He received a M.Eng. in Energy Engineering from Chiang Mai University (CMU), Thailand, in 2007. Moreover, he received a D.Eng. In Mechanical Engineering from Chiang Mai University (CMU), Thailand, in 2015. His research interests include biomass technology (gasification and pyrolysis process) and application Internal combustion Engine to biofuels. machine learning, data science, and artificial intelligence.

Tanate Chaichana received a B.Sc (Physics) from Prince of Songkla University, Thailand, in 2001. He received a M.Eng and D.Eng in the field of Energy Engineering from Chiangmai University Thailand, in 2004 and 2010 respectively. His research interests include renewable energy technologies and management for agriculture and community.

Surapon Riyana received a B.S. degree in computer science from Payap University (PYU), Chiangmai, Thailand, in 2005. Moreover, He further received a M.S. degree and a Ph.D. degree in computer engineering from Chiangmai University (CMU), Thailand, in 2012 and 2019 respectively. Currently, he is a lecturer in Smart Farming and Agricultural innovation Engineering (Continuing Program), School of Renewable Energy, Maejo University (MJU), Thailand. His research interests include data mining, databases, data models, privacy preservation, data security, databases, and the internet of things.

Tammasak Punsaensri recived a B.Eng (Agricultural Engineering) from Prince of Meajo University, Chiang Mai, Thailand, in 2002. He received a M.Eng and D.eng in the field of Energy Engineering from Chiangmai University, Chiang Mai, Thailand, in 2004 and 2019 respectively. His research interests include Agricultural Engineering, Agricultural Processes, Engineering Design, Renewable Energy, Energy Conservation, Environmental, Eco-products, Wild Fire and Natural Resource Management.