ATLANTIS PRESS

Advances in Economics, Business and Management Research, volume 91

1st International Symposium on Economic Development and Management Innovation (EDMI 2019)

Commercialization of Personal Information: Pricing and Trading

Jiaqi Chu [a], Weiping Xie [b]

North China Electric Power University, Hebei 071000, China

[a]1208537450@qq.com, [b]nanningxwp@163.com

Abstract. Nowadays, private information is becoming more and more important to both individuals and others. This private information is gradually regarded as a saleable commodity. In this paper, we construct a series of models to complete the whole task, and put forward practical policy recommendations. First, we establish an ideal framework of arbitrage free pricing model as the basic model structure of our pricing. Second, we introduce personalized differences, which define the characteristics of different individual privacy. Finally, we analyze and simulate the stability of the parameters and models, and prove that our models and methods have value. According to the establishment of the above model and the analysis of the data, the correlation between the data is found, and the privacy price pricing model is obtained.

Keywords: privacy analysis, equilibrium pricing; arbitrage free pricing principle; personalized differences.

1. Introduction

With the popularity of electronic communications and social media, some people seem willing to share personal information about their relationships and relationships (PI). Risks may include security losses, pecuniary losses, valuables losses, intellectual property losses or personal electronic property loss status. Other risks include loss of position or job, social loss, social stigmatization or marginalization. The commodity value of privacy comes from the personal value of privacy. Because privacy exists as a non-material form of information or data. It can provide personal information, can be used to understand needs, analyze more personal privacy to obtain a lot of meaningful information, personal information has certain requirements. Therefore, privacy can provide data sources and have the most basic commodity exchange principle, which can be equivalent exchanged.

1.1 Definitions and Parameters

Notation	Description
u_i, b_j	Data owner *i*, Data buyer *j*
x_i	Data element of u_i
$\hat{\varepsilon}_i$	Maximum tolerable privacy loss of u_i
w_i	Payment scheme of u_i
ε_i	Actual privacy loss of u_i in query computation
$w_i(\varepsilon_i)$	Compensation to u_i for losing ε_i
x	Data set consisting of a data element of all u_i
Q	Linear aggregated query requested by the buyer
W_{max}	Maximum budget of the buyer
W_p, W_r	Query price, Remaining budget of buyer
Q(x)	True query answer
P(Q(x))	Perturbed query answer(noise)
RMSE	Root mean squared error
X	Market maker's profit
W_{ab}	Available budget for query computation
RS	A representative sample of data set x
h	Number of representative samples RS
Φ	Number of perturbation run times

2. Model A: Personalized Differential Privacy

2.1 Definition and Analysis and Calculation

To ensure differentiated privacy, you can also learn useful information about all people when others do not know personal information. Given the privacy parameter ε, whether there is or does not have a privacy parameter, then the difference privacy level is satisfied. Whether or not there is a random disturbance result in the data set. ε can meet the needs of privacy protection, but the query results are not accurate, so we have to weigh the privacy and the accuracy of the results. In our framework, we define ε as the quantification of owner's privacy loss, which is related to price.

Definition 1. $\varepsilon-$ Differential privacy random algorithm.

M: D → R satisfies $\varepsilon-$ differential privacy if the neighboring data set a whole data set and data set *x*, *y* differs by only record, and any set of $S \subseteq \text{Range}(M)$.

$$\Pr(M(x) \in S) \leq \exp(\varepsilon) * \Pr(M(y) \in S) \tag{1}$$

Differential privacy (DP), privacy protection is for person, which means that all users in the data set have the same privacy protection / loss value. However, in practice, as our findings show, individuals may have different privacy attitudes, so allowing privacy is considered essential, especially in trading environments. Therefore, we use personalized differential privacy (PDP) theory. Each owner can personalize his or her maximum tolerable privacy level / loss, so privacy must be guaranteed to each owner. Therefore, this theory allows privacy owners to personalize privacy while providing more information.

Definition 2. $\hat{\varepsilon}$-Personalized Differential Privacy

Regarding the maximum tolerable privacy loss U, a Randomized mechanism

M:D →R satisfies every pair of neighboring data sets for any set of S ⊆ Range(M).

$$\Pr(M(x) \in S) \leq \exp\left(\hat{\varepsilon}\right) * \Pr(M(y)) \in S \tag{2}$$

Definition 3. Score Function[1]

The function f ≤ R and the probability which is proportional to the differential privacy probability of exponential mechanism output r ∈ range $(\text{F})\, s(D, r)$ is a real score function. The higher the score, the better r relative to $f(D)$. Suppose D and D' are different only on the value of the tuple, represented as $D \oplus D'$

$$s(D, r) = \max - \left| D \oplus D' \right| \tag{3}$$
$$f(D') = r$$

In personalized differential privacy, each record or data owner has its own privacy setting ε_1. To change a specific value to output. Formalize this mechanism.

Definition 4. $P\varepsilon$ principle

Give f: D → R, arbitrary input data set D ⊂ D, privacy specification $\emptyset$, mechanism $\text{P}\varepsilon_{\emptyset}^{f}(D)$ output r ∈ R probability.

$$\Pr\ [\text{P}\varepsilon_{\emptyset}^{f}(D)=\text{r}] = \frac{\exp(\frac{1}{2} d_{f(D,R,\emptyset)})}{\sum_{q \in R} \exp(\frac{1}{2} d_{f(D,r,\emptyset)})} \tag{4}$$
$$d_f(D, r, \emptyset) = max \sum i \in D \oplus D' - \emptyset^{i_u}$$
$$f(D') = r$$

Maximum tolerable privacy loss ε_{i} for all data owners in reference data x. This $\text{P}-\varepsilon$ mechanism is used to ensure that the privacy of each data owner is protected.

3. Model B: Trading Framework

3.1 Payment Scheme

In private transactions, the seller's goal as a data provider must be to maximize benefits. However, profit-driven may provide false information, but the information will still be rewarded, disrupt the normal order of the transaction, and even lead to abnormal transaction amount. Therefore, we propose a reasonable payment model, that is, to encourage information providers to provide real information and buyers to pay again. We stipulate that the payment scheme is a non-decreasing function *w: ε r*, which means that between the market manager and the data owner, the market manager should actually compensate the data owner for the lost privacy loss.

3.2 General Trading Framework

3.2.1 Constituent

1) Agency: first of all, it is the intermediary between data buyers and data owners, with a certain coordination role. The second is the appropriate way to provide information. The pricing mechanism of our country has the function of providing choice and help in theory.2) Data owners: they sell data elements by choosing the maximum tolerable privacy loss and payment scheme.3)Buyers: they purchase aggregated query answers from agents through designated queries and maximum budgets.

3.2.2 Implementation

The data owner sells his / her data element xi, to require that the actual privacy loss not exceed the prescribed maximum tolerable privacy loss. T should correspond to the selected payment plan. These data elements are stored by trusted mediation. Before the transaction, the data buyer makes a purchase request.

4. Model C: Pricing

4.1 Arbitrage Free Basic Pricing Model [2]

Pricing function, $\pi(Q)$ is a arbitrage-free equation. If every parameter S=Q_1,Q_m and Q Can be defined by S., S→Q, reach a conclusion:

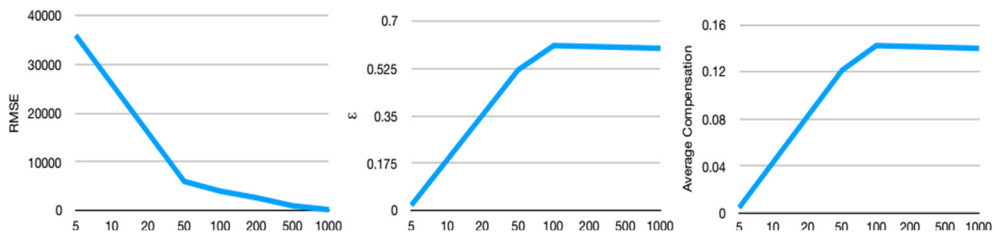
$$\pi(Q) \le \sum_{i=1}^{m} \pi(Q_i) \tag{5}$$

4.2 Simplify the Query Process

According to the principle of no arbitrage pricing, the simplified query process is as follows: Buyers must not request the same query many times, because each data owner as its own maximum privacy loss, so we must ensure that their privacy loss does not exceed their maximum tolerance. Or, the intermediary can order the query set requested by the buyer to prevent arbitrage problems.

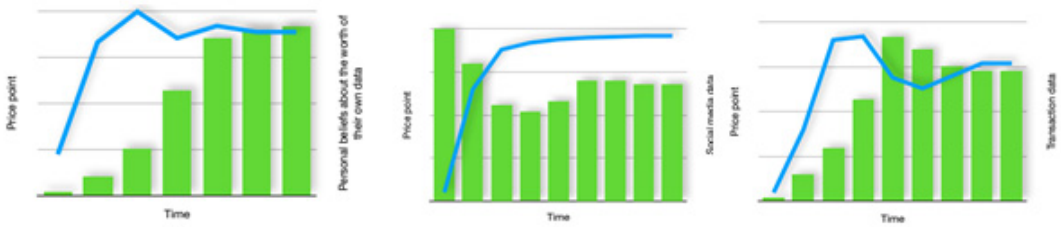
4.3 Computing Result

We extend the initial data and draw the image with the RMSE value as the dependent variable.



(a)Query price and RMSE (b)Query price and $\bar{\varepsilon}$ (c) Query price and Average compensation

Fig 1. Draw the image with the RMSE value



Personal data Social media Transaction data

Fig 2. The relationship between query price and several other factors

We simulate our mechanism to illustrate the correlation between query price and RMSE, query price and average privacy loss, and between query price and average compensation.

According to the images, we find the value of the above information shows an upward trend, and finally tends to be stable with the passage of time, indicating that the market is an accompanying relationship in t. In the early stage of information entering the market, the value of information mainly affects the market. The information increases gradually, the market forms gradually, finally the market and the information value have become a kind of mutual restriction relation, has formed a stable situation.

5. Summary

We established three model that establish a personalized privacy model to find out the correlation between risk and loss when evaluating privacy prices. We establish a trading framework, make payment plans. Thirdly, the pricing model is established, which mainly uses non-arbitrage model and equilibrium price model to determine the price.

References

[1]. Chao Li, Daniel Yang Li, Gerome Miklau, and Dan Suciu. 2014. A Theory of Pricing Private Data. ACM Transactions on Database Systems 39, 4 (2014), 1–28.

[2]. Rachana Nget, Masatoshi Yoshikawa, Yang Cao 2018 How to Balance Privacy and Money through Pricing Mechanism in Personal Data Market.