Research on Big Data Information Service Pricing Based on Uncertainty Analysis

Qian ZHANG\textsuperscript{1,a,}\textsuperscript{*} and Li-ping SHAO\textsuperscript{1,b}

\textsuperscript{1}School of Economics and Management, Beijing Jiaotong University, Beijing, China

\textsuperscript{a}18120627@bjtu.edu.cn, \textsuperscript{b}1047680753@qq.com

\*Corresponding author

Keywords: Big Data, Information Service Evaluation, Data Assets, Information Uncertainty.

Abstract. In the era of big data, more and more people use big data technology to provide users with massive data or provide information services based on the results of processing and processing massive data to meet the needs of users for specific information in specific fields. However, with the realization of internal benefits and external effects of big data, the uncertainty of big data information service and the difficulty of quality evaluation make the pricing decision and pricing game of the demander and supplier more complex and difficult to model. Based on the uncertainty of decision matching degree and service quality of big data information service, this paper analyzes the cost-benefit of data asset pricing model through index extraction and mechanism hypothesis definition, and extracts the key factors of data asset service pricing. Through the modeling of return on investment, the important influence factors in the pricing model are analyzed. From the perspective of quality evaluation, it is proposed that quality control and consumer satisfaction need to be balanced by adjusting the benefit function. That is, through the uncertainty analysis of influencing factors and price modeling analysis, the pricing research of two different business characteristics of big data information service is carried out.

1. Introduction

With the development of the era of big data, the role of big data in providing valuable information services to the outside world is gradually highlighted. Information storage and analysis methods have changed, a large number of data information to provide users with low-cost and high-efficiency information content, such as online shopping commodity push, personal business processing recommendation and other new multimedia intensive data models emerge as the times require. IBM has proposed the 5V features of big data: Volume, Velocity, Variety, Value and Veracity. Big data can be regarded as an extremely special commodity, and its use value can only be reflected after being processed by special software tools, which is also the precondition for big data to be traded.

Big data information services mainly include two types. The one is to provide services directly to users with massive data, that is, data asset services; and the other is to provide services to users based on the results of massive data analysis and mining, that is, solution services. The uncertainty analysis of big data information service pricing refers to the analysis and estimation of the change and influence of various uncertain factors on the pricing decision. Through this analysis, we can try our best to clarify and reduce the impact of uncertain factors on revenue, and provide the basis for scientific pricing and revenue maximization within a reasonable range.
2. Journals Reviewed

2.1. Research Status of Information Service Quality Evaluation

For the quality evaluation of information service, traditional methods include input evaluation method, professional standard evaluation method, task evaluation method, etc. The most important direction is the application of "user-centered" service quality evaluation theory. SERVQUAL model\[^1\] was proposed by Parasuraman et al. Its theoretical core is "service quality gap analysis model":

\[
\text{service quality} = \text{service satisfaction} = \text{service provided} - \text{expected service} = \text{perceived service} - \text{expected service}. \]

Hu C. et al.\[^2\] combined with the specific situation of our country, using the currently widely applicable analytic hierarchy process (ANP), reflected the basic content of ISO related information service in the target layer, and formed a detailed index system. On the basis of social survey, Jiao Y. and Lei X.\[^3\] proposed and verified a user satisfaction evaluation model for information resource websites from the perspective of customer satisfaction. For the quality evaluation of big data information service, because big data information service is domain specific, it is often used to provide data analysis service support and decision support for a certain field. Therefore, the quality evaluation of big data information service needs to be modeled in combination with the quality standards of specific fields. However, there are few related researches at home and abroad, which is a difficult problem in the industry. For example, in the field of cloud computing, the traditional weighted evaluation method based on quality of service (QoS) cannot dynamically evaluate the effectiveness and accuracy of large-scale cloud computing service resources. Yang X. et al.\[^4\] proposed a screening weighted evaluation method based on game optimization scheduling. This method introduces user experience quality (QoE) evaluation index to support the development of big data service business. Ma Y. et al.\[^5\] established specific indicators from three aspects of government departments, respondents and institutional factors in the government big data quality evaluation system, and completed the correction and improvement of the evaluation system in this professional field through principal component analysis.

2.2. Research Status of Big Data Information Service Pricing

The basic pricing methods of information service products mainly include cost method, market comparison method and income present value method. The cost method needs to refer to historical information and cost, which is used to evaluate trade secrets. The market comparison pricing method obtains the reference price by comparing the goods to be traded with the commodities that have been traded. This method is more direct but has poor operability because it is necessary to ensure that the reference materials are exactly the same. The income present value method is to evaluate the difference between the expected income after using big data information service and the income status before using it, and take the present value of income converted from the future cash flow brought by relevant information products as the evaluation basis. In addition to the basic pricing strategy, big data pricing strategy mainly includes static pricing strategy and dynamic pricing strategy. Static pricing strategy includes multiple pricing, discriminatory pricing, bundling pricing and Ramsey price\[^6\]. Dynamic pricing strategies include negotiation pricing, auction pricing and reverse auction pricing\[^7\]. Fu E. et al.\[^8\] selected the dynamic differential pricing method of revenue management as the main research method according to the cost particularity and difference characteristics of information goods, and constructed a dynamic pricing master model and a quantitative sub model on the relationship between information goods and consumer demand. Xiang L. et al.\[^9\] analyzed the problems in the practical application of the hierarchical pricing model, and predicted the development of the hierarchical pricing model. Starting from the characteristics of
digital content products and market environment, Cao M. et al.\textsuperscript{10} and others analyzed the relevant factors affecting pricing from two aspects of cost and network utility, and proposed to explore reasonable pricing strategies from three aspects of price orientation, price guidance and price discrimination, which are generally applicable to the pricing of digital content products. Li M. and Tian D.\textsuperscript{11} first described the lower limit and upper limit of information product price, which were the production cost (including information collection, processing, analysis and transmission process costs, information flow costs, reasonable taxes and profits) and the utility of information products to consumers (the degree of psychological satisfaction).

3. Uncertainty Analysis

The uncertainty of big data information service transaction is also an important factor affecting pricing. It is difficult to produce big data transaction, product replication is simple, product pricing is difficult and lack of reasonable trading mechanism. The transaction characteristics of these big data make the pricing problem a difficult problem in the industry. This chapter mainly combs the uncertainty of big data information services. In order to make correct pricing decisions, it is necessary to analyze these uncertain factors, calculate their impact on the decision-making scheme, and select the one with the best (or satisfactory) utility. This paper mainly studies the uncertainty of matching degree between big data information service and buyers’ decision and the uncertainty of data service quality.

(1) The uncertainty of matching degree of decision (buyers’ intention). Buyers of big data information services often need to support them to complete a strategic decision through relevant services. The matching degree between the services provided by information owners and the decision types of potential consumers effectively measures the willingness of potential consumers to purchase services. Whether consumers are willing to buy relevant big data information services depends on the environment of consumers and whether the decision-making goal matches the value of the service itself. Therefore, the transaction between big data information service providers (sellers) and service consumers (buyers) is transformed into a bilateral matching decision-making problem, which involves two different entities. Each entity gives the preference of the other party and completes the pricing decision through the index preference data. Bilateral matching decision-making problem has a wide range of practical background in the field of economic management, such as the matching problem between advertisers and readers, the matching problem between lawyers and law firms, etc. Multi index matching decision-making is a widely existing type of decision-making. It refers to how to make matching decision according to the preference of the other party under each index given by both parties. The buyer and the seller give each other the index requirements and corresponding preferences for the relevant big data services. According to the multi index matching decision-making model, the buyer can complete the purchase intention ranking of big data information service products. For the products with strong purchase intention, the buyers’ expected price is high.

(2) The uncertainty of service quality evaluation. It is assumed that in the case of high matching degree between big data information service and buyers’ intention, whether the sellers’ service can support the buyer to complete better decision is still uncertain. This uncertainty depends on the quality of the sellers’ service. The sellers’ services are mainly divided into two categories: one is the direct service of massive data assets, the other is the service of data analysis scheme.
For the scenario that massive data directly provides services to users, this paper transforms the uncertain data quality into index evaluation, as shown in Table 1.

Table 1. Definition of Data Quality Evaluation Index.

<table>
<thead>
<tr>
<th>Data quality indicators</th>
<th>Index description</th>
</tr>
</thead>
<tbody>
<tr>
<td>sample size</td>
<td>The larger the sample size, the closer to the whole sample, the higher the value of big data products</td>
</tr>
<tr>
<td>type</td>
<td>It contains a variety of data types, and different types of data have different values</td>
</tr>
<tr>
<td>Integrity</td>
<td>The higher the integrity of data, the higher the value of data</td>
</tr>
<tr>
<td>time span</td>
<td>The longer the time span, the higher the value</td>
</tr>
<tr>
<td>Real time</td>
<td>Real time data is more valuable than historical data</td>
</tr>
<tr>
<td>depth</td>
<td>For a certain type of data, the more thorough the analysis, the higher the value of the data</td>
</tr>
<tr>
<td>Sample coverage</td>
<td>The larger the data breadth and dimension, the higher the sample coverage and the higher the value of data products</td>
</tr>
<tr>
<td>Scarcity</td>
<td>The rarer, rarer and more valuable data is</td>
</tr>
</tbody>
</table>

This chapter completes the index extraction and problem transformation of the above uncertainty factor analysis. It is necessary to establish pricing models for different business types.

4. Analysis of Information Service Pricing

This chapter gives the practical research of information service pricing. Combined with literature review and research, this chapter established the existing mechanism of data supplier, market owner and consumer under the background of data pricing, and designs and analyzes the data asset pricing model. At the same time, this chapter explores the solution pricing model based on the characteristics of large difference in investment returns.

4.1. Research on Pricing Practice

For the two kinds of big data information service forms classified in the introduction, this chapter describes and analyzes the practical research on service pricing. The following is the practical research on several data providers, including the description of product types, pricing methods and pricing basis. As shown in Table 2.

Table 2. Information Service Pricing Practice Research.

<table>
<thead>
<tr>
<th>corporate name</th>
<th>product type</th>
<th>pricing methods</th>
<th>Pricing basis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Shop</td>
<td>Enterprise data, business data</td>
<td>Fixed pricing for all kinds of data</td>
<td>Data acquisition and processing difficulty, data volume size</td>
</tr>
</tbody>
</table>
Data Hall | Artificial intelligence data set and scheme | Data set fixed pricing | Data acquisition and processing difficulty, data volume size
--- | --- | --- | ---
Friendship+ | PC/APP/offline statistics, advertising effect monitoring, Industry Report | Free for the first time - fixed price | The difficulty of data acquisition and processing
GISSO Robot Information | Customized information | Unit price (data flow) | Customer demand, technical difficulty and data flow
BIT | BI tools, big data analysis, big data display and processing | Free Admission | —

It is not difficult to find that in the practical research of the above companies, most of them are charged. The data service pricing often takes into account the technical realization cost and the definition of the completion price of specific business indicators. Through the KPI analysis of the service target group and the enterprise itself, the pricing strategy and business development mode are completed.

4.2. Design of Data Asset Pricing Model

In the context of data pricing, stakeholders are data providers, market owners and consumers. On the data market platform, providers provide raw data from various sources to provide a certain number of data samples for potential consumers. Consumers search data sets on the platform according to their needs and preferences, and make purchase decisions based on perceived value and willingness to pay. Government agencies, companies and even individuals can be data providers and consumers. Data platform owners manipulate data obtained from multiple data providers and constantly update dynamic data. As owners of valuable data, platform owners provide trading platforms and set rules to manage transactions between providers and consumers. Like other commodities, data transactions are guided by quality, price and the "willingness to pay" of potential consumers [11]. The purpose of this paper is to solve the pricing problem from the perspective of data market owners. The purpose is to analyze the important pricing factors that data platform owners need to consider when making decisions on data quality and price.

Suppose a monopolized data market in which the monopolist has the right to make rules and provide a trading platform for data providers and consumers. Monopoly data platform owners provide different data choices, and these consumers are different in terms of different data quality preferences. Data platform is based on the maximum total profit to control decisions and guide consumers, and consumers maximize the utilization of data through their own choice. According to the display principle, in such a market environment, when all participants try to "tell the truth", an optimal strategy will be achieved. "Telling the truth" means that data providers provide high-quality and high compliance data, which can truly reflect the laws of the objective world. According to the evaluation index of data quality in Chapter 3, we can define the compliance and non-compliance of data asset quality. Through the scoring mechanism and linear weighted average, the non-compliance
score of data can be obtained, which is set as $p$ ($0 < p < 1$), the real income of compliance data in all data assets is $E_0$, and the additional income of non-compliance data is $R$.

$$E_0 \leq (1 - p) \cdot (E_0 + R)$$ (1)

In other words, the income of data provider is less than or equal to the sum of the income of compliance data and non-compliance data. In the real world, the data compliance and quality inspection discovery is based on probability, that is, if the probability of data non-compliance is defined as $p_1$, the revenue is ‘$E + R$’ when the non-compliance data is not found, and when the non-compliance phenomenon is found, the income of the data provider is determined as $E_1$ (earnings decrease due to credit loss, i.e., $E_0 < E_1$). Then (1) formula is as follows:

$$E_0 \leq (1 - p_1) \cdot (E_0 + R) + p_1 \cdot E_1$$ (2)

(2) The results are as follows:

$$E_0 - E_1 \leq \left(\frac{1}{p} - 1\right) \cdot R$$ (3)

From the above formula, the left side of the inequality is the loss of revenue caused by the non-compliance data, and the right side is the potential income brought by the non-compliance data. According to the above formula (2) (3), if the extra income of the non-conforming data is greater, the probability of being found is smaller, then the expected return will be larger. Therefore, in this market environment, its characteristics are similar to the market in which bad money drives out good money. Therefore, in this chapter, in addition to data providers and participants, it is necessary for the monopoly market owners to include the credit score of data providers into the pricing function when defining the pricing function of data information services. Once the non-compliance data provided by the data provider exceeds a certain threshold, the product pricing will be reduced for its credit. In order to avoid the phenomenon of revenue expansion of non-compliance data, and support the market owners to complete the final information service pricing.

### 4.3. Design of Solution Pricing Model

In the context of data solution pricing, the team providing big data analysis solutions is often both a service provider and a pricing participant. In this case, the cost plus method is highly interpretable, as described in Chapter 3 for solution pricing. However, considering that the value of data service is an investment in the future, because the knowledge of the data service scheme has a life cycle, the transaction also has a life cycle. The higher the rate of return, the higher the depreciation rate. Therefore, when considering the solution pricing, we need to introduce the concept of return on investment as the basis to establish the price model.

Let $r$ be the rate of return on investment, $Q$ be the volume of service transactions (the number of service buyers), and $C$ be the development cost, including data acquisition cost $C_r$ and data analysis cost $C_s$. For solution services, it is assumed that the data acquisition cost $C_r$ does not change with $Q$.
(the fixed Cr is calculated by using the cost pricing method to describe the working hours), while the data acquisition cost Cs changes little and can be regarded as fixed. EV is expected return. There are:

\[ C + EV = P*Q \]  

(4)

Because the expected return can be expressed as the product of the investment yield r and the development cost C:

\[ EV = r*C \]  

(5)

Therefore, by substituting formula (5) into equation (4), we can get the following results:

\[ P = \frac{C + EV}{Q} = \frac{(1+r)C}{Q} \]  

(6)

In other words, the pricing of solutions is related to r, Q, C, while the return on investment is closely related to the transaction volume, service quality and sustainability of services. In addition, there are significant differences in the rate of return of different industries, such as consulting industry and education industry. That is, the rate of return on investment has a significant impact on the scientific pricing of decision-making schemes, but the relevant research is industry-related, which is not the focus of this paper.

5. Quality Assessment Strategy

For the quality evaluation of data asset services, this paper discusses in 4.2 that the quality of data assets can be measured from multi dimensions by weighted scores for data indicators. However, the quality evaluation of data solution services is often related to customer satisfaction to a large extent. This paper argues that the more satisfied the customer is, the higher the profit of related information services is. If the information service provider changes the information service plan frequently in the later stage of service in order to meet the needs of customers, the quality improvement fee paid by relevant teams is extremely high. Therefore, we should grasp the customer satisfaction and balance the revenue and expenditure on this basis. Reaching the dotted line in Figure 1 is an advantageous strategy for service providers and service consumers. Among them, the quantity improvement measure cost is QIC, the quality management cost is C2, the internal loss cost and external loss cost are C1 and C3.
Therefore, to sum up, the establishment of a comprehensive evaluation strategy for data quality and service quality, and according to 4.2, give appropriate punishment measures to market participants from the perspective of credit system can effectively ensure the integrity and correctness of service quality, so as to reduce the uncertainty of big data information service.

6. Summary and Prospect

This paper summarizes the research status of big data information service quality evaluation and big data information service research status, completes the uncertainty analysis of big data information service, and takes the decision matching degree uncertainty and service quality uncertainty as the main analysis objects, models the uncertainty factors, extracts indicators, and analyzes the data asset pricing model through problem transformation Design and solution pricing model design. This paper points out that in the market mechanism of data providers, market providers and data consumers, the role of credit score in achieving the global equilibrium of data asset income is clarified. In the market mechanism of service providers and service consumers, the evaluation of return rate is introduced to complete the pricing of the solution.

The pricing of big data information service has strong domain relevance. This paper obtains the main factors influencing the pricing and game analysis through macro analysis. However, it is the future research focus to improve and expand the pricing model combined with the quality evaluation system in related fields.

References


