Research Article

Discrete Hyperparameter Optimization Model Based on Skewed Distribution

Yuqi Li

School of Economics and Management, Beijing University of Posts and Telecommunications, Beijing 100876, China

Correspondence should be addressed to Yuqi Li; liyuqi@bupt.edu.cn

Received 21 March 2022; Revised 15 June 2022; Accepted 21 June 2022; Published 9 August 2022

Academic Editor: Man Fai Leung

Copyright © 2022 Yuqi Li. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

As for the machine learning algorithm, one of the main factors restricting its further large-scale application is the value of hyperparameter. Therefore, researchers have done a lot of original numerical optimization algorithms to ensure the validity of hyperparameter selection. Based on previous studies, this study innovatively puts forward a model generated using skewed distribution (gamma distribution) as hyperparameter fitting and combines the Bayesian estimation method and Gauss hypergeometric function to propose a mathematically optimal solution for discrete hyperparameter selection. The results show that under strict mathematical conditions, the value of discrete hyperparameters can be given a reasonable expected value. This heuristic parameter adjustment method based on prior conditions can improve the accuracy of some traditional models in experiments and then improve the application value of models. At the same time, through the empirical study of relevant datasets, the effectiveness of the parameter adjustment strategy proposed in this study is further proved.

1. Introduction

The whole Pareto/NBD model was created by Abramowitz and Stegun. The practical problem corresponding to the initial modeling is to analyze the behavior of consumers, especially the repeated purchase of related products, so that managers can evaluate the purchasing intention of future consumers and better arrange production and marketing activities. For the whole model itself, the description of consumer behavior is reflected in the following two important parts [1]. First, it is about the probability distribution of the consumer’s survival over the entire product purchasing cycle; second, it is about the description of the mathematical expectation of consumers’ purchasing behavior randomly selected in a certain consumption period in the future.

For the construction of the whole Pareto/NBD model, the most critical part is to get the exact mathematical analytical expression of conditional mathematical expectation through appropriate mathematical assumptions and strict mathematical reasoning [2–4]. This work is very difficult for many researchers who are familiar with the SMC model. An important objective of this study was to prove the accuracy of some key conclusions about the Pareto/NBD model through mathematically deriving relevant important intermediate results (especially probability distribution function and probability distribution density function) within the framework of Pareto/NBD model. At the same time, the analytical framework of this model is combined with the hyperparameter optimization problem of discrete SMBO type to achieve the accuracy of mathematical description of the whole model.

First of all, we need to introduce a special function, Gauss hypergeometric function, and the mathematical expression of this function in the form of power series [5] is as follows:

\[ \sum_{j=0}^{\infty} \frac{(a)_j (b)_j}{(c)_j j!} z^j \]  

(1)

where \( c \neq 0, -1, -2, \ldots \), \( (a)_j \) is the factorial power (Pochhammer’s symbol) operation, and specific mathematical calculation formula is \( a(a+1) \cdots (a+j-1) \). In contrast to the forming operation, it is about the parameter \( a \) of the
ascending factorial power operation, and according to the assessment method of Abel, the convergence of the factorial power operation depends on the parameter $z$, in the case of $|z| < 1$ throughout the series converges to 1, in the case of $|z| > 1$ tends to spread the whole series, $|z| = 1$, the series converges to the $c - a - b > 0$.

The ascending joint operation $(a)_j$ can be expressed by two gamma functions with related parameters, which can be expressed by mathematical formula as follows:

\[
(a)_j = \frac{\Gamma(a+j)}{\Gamma(a)}
\]

(2)

With the help of this nature, the mathematical expression of the whole Gauss hyperparametric function can be further expressed as follows:

\[
_2F_1(a, b; c; z) = \frac{\Gamma(c)}{\Gamma(a)\Gamma(b)} \sum_{j=0}^{\infty} \frac{\Gamma(a+j)\Gamma(b+j)}{\Gamma(c+j)} \frac{z^j}{j!}.
\]

(3)

According to the above mathematical expression form of Gauss hypergeometric function, combined with the nature of gamma function, the whole function is still expressed in discrete form, and the whole Gauss hypergeometric function is mutually symmetric for two parameters $a$ and $b$ [6], which is expressed in mathematical form as follows:

\[
_2F_1(a, b; a; z) = \frac{\Gamma(c)}{\Gamma(a)\Gamma(b)} \sum_{j=0}^{\infty} \frac{\Gamma(a+j)\Gamma(b+j)}{\Gamma(c+j)} \frac{z^j}{j!}.
\]

(4)

Furthermore, to further deepen the understanding of Gauss hypergeometric functions and facilitate the next model derivation, we must introduce the representation of Gauss hypergeometric functions in the form of Euler integral:

\[
_2F_1(a, b; c; z) = \frac{1}{B(b, c - b)} \int_0^1 (1-t)^{b-1}(1-zt)^{-a} dt, c > b.
\]

(5)

Through the above mathematical expressions, we get the continuous expression of the Gauss hypergeometric function, where $B(\cdot, \cdot)$ is the beta function.

2. Model Assumptions

Many aspects of BG/NBD model are similar to the Pareto/NBD model, but there are still many differences between models. The most important part is whether the hyperparameters change with each experiment/test. In the Pareto model, the hyperparameters change in any interval of each experiment/test interval, which is further independent of the selection and change in datasets. In the BG/NBD model, we assume that the timing of determining whether the hyperparameters change is controlled after each new dataset test; that is, the continuous discrimination interval in Pareto/NBD is changed into discrete decision interval, and then, the beta-geometric (BG) model can be established [7]. On the whole, the following five basic assumptions are needed for the new model.

(1) When dataset testing further influences the change in hyperparameters, the change in hyperparameters is subject to the Poisson process with parameter $\lambda$ as a whole; that is to say,

\[
P(\Theta(d_i) = x|\lambda) = \frac{(\lambda d_i)^x \exp(-\lambda d_i)}{x!},
\]

(6)

where $d_i = |D_i - \emptyset|$ represents the difference between different datasets and empty sets, which can be the complexity change between datasets or the size change in datasets. For Poisson process distribution, the exponential distribution with parameter $\lambda$ is followed between every two dataset experiments, and the mathematical expression is as follows:

\[
g(d_j - d_{j-1}| \lambda) = \lambda \exp[-\lambda(d_j - d_{j-1})], \quad d_j > d_{j-1} \geq 0,
\]

(7)

where $g$ stands for probability density function.

(2) For parameter $\lambda$, we assume that it is subject to the gamma distribution of parameters $r$ and $\alpha$, and the mathematical expression is as follows:

\[
g(\lambda|r, \alpha) = \frac{\alpha^r \lambda^{r-1} \exp(-\lambda)}{\Gamma(r)}, \quad \lambda > 0,
\]

(8)

where $r$ and $\alpha$ are the known and determined constants, which are determined by exogenous factors and not affected by the model itself.

(3) After each experiment/test, the probability $p$ of the hyperparameter $\theta$ will remain unchanged. In other words, after any experiment/test, a judgment should be made on whether the next hyperparameter $\theta$ will change further. Then, the probability of the hyperparameter $\theta$ remaining unchanged should follow the geometric distribution of the parameter $p$:

\[
P(\text{unchange}) = p(1-p)^{j-1},
\]

(9)

where $j$ is the number of experiments/tests.

(4) To ensure that there is an analytic solution to the posterior distribution involving hyperparameters when solving the Bayesian estimation, it is necessary to assume that the parameter $p$ is also determined by the exogenously given probability distribution, namely,

\[
g(p|a, b) = \frac{p^{a-1}(1-p)^{b-1}}{B(a, b)}, \quad 0 \leq p \leq 1,
\]

(10)

where $a$ and $b$ are the known and determined constants, which are determined by exogenous factors and not affected by the model itself.

(5) To eliminate the interaction between parameters and simplify the complexity of the overall model analysis, we need to assume the independence of parameters $\lambda$
and $p$ of the two models; that is, with the progress of the experiment/test, $\lambda$ and $p$ are independent of each other and not affected by each other.

### 3. Model Evolution

3.1. Derivation of Likelihood Function. We may consider that for the hyperparameter $\theta$, it has experienced $x$ changes in the known experimental/test interval $(0,d_T]$. These changes may occur in $x$ interval $d_1,d_2,\ldots,d_x$ in the whole test interval, which can be represented by the number line graph as follows.

To ensure that a relatively complete likelihood function of relevant parameters can be derived [8], we carry out the following technical operation and further assumption:

1. The probability of hyperparameter $\theta$ changing between $d_1$ in the first experiment/test area is a standard exponential distribution model, and the probability of event occurrence should be $\lambda \exp(-\lambda d_1)$.

2. The probability that the second change in hyperparameter $\theta$ happens to be in the second experiment/test interval $d_2$ is in the first experiment/test interval $d_1$. The probability of hyperparameter $\theta$ still continues to change with the product of the probability of $d_1$ and $d_2$ changes within the range of test; that is to say, in the whole experiment/test interval, the probability of $\theta$ hyperparameter change still obeys the standard exponential distribution, but needs to multiply by the probability that the hyperparameter $\theta$ will still change in the first test interval. The specific expression is as follows:

$$
(1 - p) \lambda \exp(-\lambda (d_2 - d_1)).
$$

(11)

3. Further extend to the probability that the hyperparameter $\theta$ changes at the $x$th experiment/test interval $d_x$ should be equal to the standard exponential distribution times the probability that the change will occur after the previous experiment/test interval $d_{x-1}$, and the mathematical expression is as follows:

$$
(1 - p) \lambda \exp(-\lambda (d_x - d_{x-1})).
$$

(12)

4. Finally, to ensure the likelihood function expression of completeness, we must make mandatory assumptions about the predicted change probability after the first experiment/test $d_1$, assuming that the first test/test must be changed, but the changes can take 0 concrete or other step length, so overall hyperparameter $\theta$ in $(d_{x},d_T]$ interval constant probability is composed of two parts. The first part is the first time the experiment/test has changed, both of which do not change afterwards, and the second part is just the possibility just there is no change in $(d_{x},d_T]$ interval, where the specific expression is as follows:

$$
p + (1 - p) \lambda \exp(-\lambda (d_T - d_x)).
$$

(13)

Further, we can obtain the likelihood function of the parameter $d_1,d_2,\ldots,d_x,d_T$ when $\lambda$ and $p$ are known:

$$
L(\lambda, p|d_1,d_2,\ldots,d_x,d_T) = \lambda \exp(-\lambda d_1)(1-p) \\
\cdot \lambda \exp(-\lambda (d_2-d_1)) \ldots (1-p) \\
\cdot \lambda \exp(-\lambda (d_x-d_{x-1})) \\
\cdot \lambda \exp(-\lambda (d_T-d_x)) \\
\cdot \{p + (1-p)\exp(-\lambda (d_T-d_x))\} \\
= p(1-p)^{x-1} \lambda^x \exp(-\lambda d_x) \\
+ (1-p)^x \lambda^x \exp(-\lambda d_T).
$$

(14)

Just like our Pareto/NBD model above, the change in hyperparameter $\theta$ is still uncertain and random, and the prediction of the overall hyperparameter change depends on the historical information of the whole experiment/test: $(\Theta = x,d_x,d_T)$.

At the zero point of the initial experiment/test preparation stage, we must ensure that the hyperparameter $\theta$ will change, but there is a difference in the change step size, and then, we can get the probability likelihood function of the hyperparameter $\theta$ remaining at 0 in the whole historical information range of $(0,d_T]$, which can be further obtained using the exponential distribution:

$$
L(\lambda|\Theta = 0,d_T) = \exp(-\lambda d_T).
$$

(15)

Further, for a single hyperparameter $\theta$, we can obtain a probability likelihood function, and the mathematical expression is as follows:

$$
L(\lambda, p|\Theta = x,d_T) = (1-p)^x \lambda^x \exp(-\lambda d_T) \\
+ I_{x=0}p(1-p)^x \lambda^x \exp(-\lambda d_x),
$$

(16)

where $I_{x=0} = 1$, when $x > 0$, and $x \leq 0$ is less than minus, and the value of the whole is 0.

3.2. Derivation of the Probability Distribution of the Hyperparameter $\theta$. For $\theta$ hyperparameter probability distribution, the application of mathematical formula is expressed as $P(\Theta(d_i) = x)$, where the $\Theta(d_i)$ represents the specific value of hyperparameter $\theta$ in the whole experiment/test interval $d_i$, so $\Theta(d_i)$ becomes a random variable, so that you can establish the basic relationships between experiment/test interval and hyperparameter $\theta$: $\Theta(d_i) \geq x = T_x \leq d_i$. Among them, $T_x$ shows experiment/test range when $\theta$ hyperparameter is $x$ changed. According to the corresponding relationship, we can get approximate mathematical expressions about $P(\Theta(d_i) = x)$:

$$
P(\Theta(d_i) = x) = P(\text{change}) \cdot P(T_x \leq d_i, T_{x+1} > d_i) \\
+ I_{x=0} \cdot P(\text{unchange}) \cdot P(T_x \leq d_i).
$$

(17)

Considering that the experimental/test interval in which the hyperparameter $\theta$ changes follows an exponential distribution (for specific content, see model hypothesis part), according to random variable for the hyperparameter itself, an important part of the probability distribution function is
\( P(T_x \leq d_x, T_{x+1} > d_x) \), a Poisson probability problem; among them, \( \Theta(d_x) = x \) and \( P(T_x \leq d) \) (Erlang-\( x \)) are the arrival time distribution. Furthermore, we can derive the conditional distribution function of hyperparameter \( \theta \) with machine variable under the known condition of RP [9]:

\[
P(\Theta(d_x) = x|\lambda, p) = (1 - p)^{\lambda \theta(d_x)} \exp(-\lambda \theta(d_x)) \]

\[
\quad \quad \quad \quad \quad \quad + I_{x=0} p (1 - p)^{x-1} \]

\[
\quad \quad \quad \quad \quad \quad \cdot \left[ 1 - \exp(-\lambda \theta(d_x)) \sum_{j=0}^{\infty} \frac{(\lambda \theta(d_x))^j}{j!} \right].
\]  

(18)

3.3. Derivation of Mathematical Expectations for the Hyperparameter \( \Theta \). Since the change in the hyperparameter \( \Theta \) follows the Poisson process with the parameter \( \lambda \theta \), the mathematical expectation of the value change in the hyperparameter \( \Theta \), i.e., \( E[\Theta(d_x)] \), should be \( \lambda \theta \theta \), within the given experimental/test interval. Then, for the hyperparameter \( \Theta \), there are no longer changes in the experimental/test interval \( d_x \); then, the mathematical expectation of \( E[\Theta(d_x)] \) in \( (0, d_x) \) with respect to the hyperparameter \( \Theta \) should be equal to \( \lambda \theta \).

Then, for the change occurring outside the experimental/test interval \( d_x \), the probability distribution of the change in the value of the hyperparameter \( \Theta \) should follow the conditional probability model about \( \lambda \) and \( p \), that is,

\[
P(d_x > d) = P(\text{change}|\lambda, p)
\]

\[
\quad = \sum_{j=0}^{\infty} (1 - p)^{j-1} \exp(-\lambda \theta(d_x)) \frac{(\lambda \theta(d_x))^j}{j!}
\]

\[
\quad = \exp(-\lambda \theta d).
\]  

(19)

This expression further indicates the probability density function for no-answer experiment/test interval \( d_x \); that is to say, \( g(d_x|\lambda, p) = \lambda \exp(-\lambda \theta(d_x)) \). It shall be highlighted that no-answer experiment/test interval \( d_x \) is an exponential form, but the relevant parameter arrangement depends on the probability for the recognition of each experiment/test \( p \). However, in a Pareto/NBD model, the overall model assumes that no-answer experiment/test interval \( d_x \) is specified by the probability density function, and there is no relevant information for the discrimination of probability \( p \). Then, we can further work out the conditional expectation expression about \( \Theta(d_x) \) in the case that \( \lambda \) and \( p \) know, and the specific mathematical expression is as follows:

\[
E[\Theta(d_x)|\lambda, p] = \lambda \theta \theta \cdot P(d_x > d) + \int_{0}^{d_x} \lambda \theta \theta \cdot g(d_x|\lambda, p) d(d_x).
\]  

(20)

By substituting relevant intermediate variables into the above formula and further simplifying, we can get a very concise formula:

\[
\frac{1}{p} - \frac{1}{p} \exp(-\lambda \theta d).
\]  

(21)

3.4. Derivation of the Transcendental Posterior Expectation. By the content of the previous section, we discuss the core of the problem is to deduce the mathematical expectation of \( E[\Theta(d_x)|\lambda, p] \) involving \( \Theta \) hyperparameter under the condition of known parameters \( \lambda \) and \( p \), that is, under the premise of arrival rate \( \lambda \) of the test data and the know \( p \), the conditional expectation expression of the hyperparameter \( \theta \) is obtained, but in fact, according to the previous hypothesis, according to \( \lambda \) and \( p \), we set and select the two parameters that obey the distribution types and determined parameters. This part mainly focuses on the content of the derivation in \( \lambda \) to gamma distribution and \( p \) to beta distribution of cases, that is, to derive the posterior conditional expectation of \( \Theta(d_x) = x \) involving the parameters \( r, \alpha, \beta, b \) when we assume that the conditions (1) and (2) meet. The whole derivation process is relatively complex, and we will treat the proof content as an appendix. This summary mainly lists the main results in the derivation process.

(1) According to the content of Section 3.1, we have derived the probability likelihood function of the related parameters \( \lambda \) and \( p \) when the hyperparameter and the whole experimental/test interval are known, in the form of formula (3). Furthermore, we will make parameters \( r, \alpha, \beta \) obey the parameters for the gamma distribution and \( a, b \), respectively, the assumptions for the beta distribution into the middle of the derived formula; note that we need to emphasize that the entire \( (r, \alpha, a, b) \) parameter set is given and then get in the whole experiment/test cycle (\( \Theta = x, d_x, d_T \)) known cases, and related parameter \( (\Theta = x, d_x, d_T) \) probabilistic likelihood function is expressed as follows:

\[
L(r, a, b|\Theta = x, d_x, d_T) = \frac{B(a, b + x) \Gamma(r + x + a')}{B(a, b) \Gamma(r)\Gamma(a + d_x)}[1 - \frac{B(a + 1, b + x - 1)}{B(a, b) \Gamma(r)(a + d_x)}]
\]

(22)

All the four variables \( (r, a, b, \alpha, \beta) \) in the above formula can be estimated by means of maximum-likelihood estimation. That is to say, in the whole known experiment/test cycle \( (0, d_T) \), for \( \Theta = x \), the probability likelihood function after the log operation occurred in the \( d_x \) experiment/test interval can be expressed as follows:

\[
LL(r, a, b|\Theta = x, d_x, d_T) = \ln[L(r, a, b|\Theta = x, d_x, d_T)].
\]  

(23)

The estimation of this formula can follow the general numerical optimization method and determine the optimal solution by finding the stagnation point of the first derivative and solving the extreme value, because the nature of the likelihood function must be convex.
(2) According to the conditional probability distribution of hyperparameter $\theta$, the form is shown in formula (4). Considering that the specific distribution of $\lambda, p$ has been given and the parameters have been determined, we can further express formula (4) as the conditional probability distribution under the condition that parameter set $(r, a, a, b)$ is known:

$$P(\Theta|d_i) = \frac{B(a + b + x)}{B(a,b)} \frac{\Gamma(r + x)}{\Gamma(r)x!} \left(\frac{d_i - 1}{\alpha + d_i}\right)^x$$

$$= \frac{B(a + 1, b + x - 1)}{B(a,b)} \left[1 - \left(\frac{\alpha}{\alpha + d_i}\right)^r \sum_{j=0}^{\infty} \frac{\Gamma(r + j)}{\Gamma(r)j!} \left(\frac{d_i - 1}{\alpha + d_i}\right)^j\right].$$

(24)

(3) Finally, according to the condition of known parameters of $\lambda, p$, the hyperparameter $\theta$ conditional expectation, the concrete mathematical expression such as formula (5), combined with the $\lambda, p$ two parameters, follows gamma distribution and beta distribution in concrete form, and we can get in any experiment/test interval length $d_i$, under the condition of the parameters in the parameter set $(r, a, a, b)$ known, hyperparameter $\theta$ conditional expectation, which can also be called the posterior conditional expectation; the mathematic expression of the specific is as follows:

$$E[\Theta|d_i|r, a, a, b] = \frac{a + b - 1}{a - 1}$$

$$\cdot \left[1 - \left(\frac{\alpha}{\alpha + d_i}\right)^r \sum_{j=0}^{\infty} \frac{\Gamma(r + j)}{\Gamma(r)j!} \left(\frac{d_i - 1}{\alpha + d_i}\right)^j\right].$$

(25)

where $F_1(\cdot)$ is Gauss hypergeometric function; refer to the explanation in the proof for the simple derivation, and refer to the relevant introduction in the model introduction for the specific content.

It is important to note that the final expression of posterior conditional expectation about hyperparameter $\theta$ needs a value about Gauss hypergeometric function; of course, this value has nothing to do with hyperparameter $\theta$. On the contrary, due to the parameter set $(r, a, a, b)$ can be obtained by maximum-likelihood estimate of the traditional method, and the experiment/test interval length is also a known quantity, so the overall Gauss hypergeometric function value is also determined. Compared with other estimated results, using the Gauss hypergeometric function to express the hyperparameter $\theta$ of the posterior conditional expectation is more direct and simple, and polynomial sequence can be used for approximate calculation, so the convenience for the later of simulation is very good, which can greatly simplify the computing complexity and try to solve the corresponding parameter selection in a timely manner.

Of course, the above content is just based on known sequence information $(\Theta = x, d_x, d_F)$ cases and makes the probability model for how hyperparameter $\theta$ values. In addition to accommodating more known conditions and historical information as much as possible to make accurate simulation of the structure of past data, the final purpose of establishing BP/NBD model shall also focus on a given new experiment/test dataset case and how to utilize the known datasets and the relationship between the parameter $\theta$, given the new dataset the parameters of an approximate estimate [10], and the key of the whole model application on the new dataset corresponds to the hyperparameter of the prediction is that we more concern.

$(\Theta = x, d_x, d_F)$ in the history of the experiment/test information has been confirmed, we can have $\theta$ hyperparameter values on the new experiment/test interval $d_f$, we give an approximate posterior conditional expectation, the specific detailed process can see behind a certificate, and here we give the concrete mathematical expression of this posterior conditional expectation:

$$E[Y(d_f)|\Theta = x, d_x, d_F, r, a, a, b]$$

$$= \frac{a + b + x - 1/a - 1}{1 + I_{\text{aff}}a/b + x - 1(a + d_f/a + d_f)}$$

$$\cdot \left[1 - \left(\frac{\alpha}{\alpha + d_i}\right)^r \sum_{j=0}^{\infty} \frac{\Gamma(r + j)}{\Gamma(r)j!} \left(\frac{d_i - 1}{\alpha + d_i}\right)^j\right].$$

(26)

where $\Theta = x, d_x, d_F$ stands for the known experimental/test information, $(r, a, a, b)$ stands for the known parameter set by calculation, and $F_1(\cdot)$ stands for the Gauss hypergeometric function.

Once again, the whole formula uses the Gauss hypergeometric function, and a mathematical expression formula is the same, of course, where the Gauss hypergeometric function parameters are known, and the calculation of the precise value is guaranteed. To reduce the computation burden on the computer simulation process, the actual operation often uses polynomial equations to approximate its numerical calculation and will not cause enormous computation burden. The rest of the whole formulas are simple numerical calculations and would not form the computing burden.

The overall content of the appendix mainly focuses on the mathematical description of the key process in the model derivation process. On the one hand, through mathematical...
rigour, it indicates the credibility of model content; on the other hand, it also provides mathematical support for transforming the mathematical model into computer model. The main purpose is the mathematical expectation for hyperparameter \( \theta \), \( E[\Theta (d)] \), and ultimately predictable \( \theta \) hyperparameter posterior mathematical expectation of mathematical deduction, and the derivation process of the intermediate links is the Euler integral for Gauss hypergeometric function:

\[
\int_{0}^{1} \binom{a}{b} \binom{c}{d} \leq (1-t)^{-a-b} (1-zt)^{-c-d} \, dt, \quad c > b.
\]  

(27)

3.5. Derivation of Mathematical Expectation \( E[\Theta (d)] \) for the Hyperparameter \( \theta \). To get the mathematical expectation \( E[\Theta (d)] \) about hyperparameter \( \theta \), according to the condition of known parameter set \( \lambda, \rho \), we must use formula about the conditional expectation of hyperparameter \( \theta \) and assumptions about the distribution of \( \lambda \) and \( \rho \), \( \lambda, \rho \), parameters of \( r, a, b \) gamma distribution on \( \rho \) obedience of \( a, b \) beta distribution parameters; first of all, we will bring the probability distribution density of \( g(\lambda | r, a) = (\lambda^a r^{-1} \exp(-\lambda r)) \), \( \lambda > 0 \), related to \( \lambda \) into formula (4) and get preliminary mathematical expressions:

\[
E[\Theta (d) | r, \alpha, \beta] = \frac{1}{\beta} - \frac{\alpha^\rho}{\rho (a + \rho d)^\rho}.
\]  

(28)

Further, we substitute the beta probability distribution density \( g(p | a, b) = (p^a (1-p)^b / B(a, b)) \), \( 0 \leq p \leq 1 \), of \( p \) subject to the overall conditional expectation formula of hyperparameter \( \theta \), and we first calculate a definite integral as follows:

\[
\int_{0}^{1} \frac{p^{-a-1} (1-p)^{-b-1} + \rho | B(a, b)}{B(a, b)} \, dp = \frac{a + b - 1}{a - 1}.
\]  

(29)

Further, a more complex definite integral needs to be calculated:

\[
\int_{0}^{1} \frac{\alpha^\rho}{\rho (a + \rho d)^\rho} \, dp = \frac{\alpha}{\rho} \int_{0}^{1} p^{-a-2} (1-p)^{-b-1} (a + \rho d)^{-\rho} \, dp.
\]  

(30)

To ensure that the definite integral form at the end is consistent with the form of Gauss hypergeometric function, we need to make a variable substitution, so \( q = 1 - p \); of course, in the process of variable substitution, the form of the differential will also change \( dp = dq \). The integral form above further becomes the following:

\[
\left(\frac{a}{a + \rho d}\right)^\rho \int_{0}^{1} q^{-a-2} (1-q)^{-b-2} \left(1 - q \frac{d}{a + \rho d}\right)^{-\rho} \, dq.
\]  

(31)

The Euler integral form and parameter \( r, b; a + b - 1; (d/\rho + d/\rho) \) of the Gauss hypergeometric function mentioned above are further applied, and the corresponding parameters in the integral form are expressed, respectively; then, the definite integral can be further formalized as follows:

\[
= \left(\frac{a}{a + \rho d}\right)^\rho \int_{0}^{1} q^{-a-2} (1-q)^{-b-2} \left(1 - q \frac{d}{a + \rho d}\right)^{-\rho} \, dq.
\]  

(32)

where \( F_1(\cdot) \) is a Gauss hypergeometric function.

Finally, the conditional expectation formula of the hyperparameter \( \theta \) can be obtained, and its mathematical expression is as follows:

\[
E[\Theta (d) | r, \alpha, \beta] = \frac{a + b - 1}{a - 1} \left[ 1 - \left(\frac{a}{a + \rho d}\right)^\rho \int_{0}^{1} q^{-a-2} (1-q)^{-b-2} \left(1 - q \frac{d}{a + \rho d}\right)^{-\rho} \, dq \right].
\]  

(33)

3.6. Derivation of the Conditional Expectation \( E[Y (d)] \) for the Hyperparameter \( \theta \) with Known Historical Information. We further define the variation of \( (d/\rho, d/\rho, d/\rho) \) hyperparameter \( \theta \) in the new experimental/test interval of \( Y (d) \). As a prediction of hyperparameter \( \theta \) in the new dataset, we focus on the conditional expectation of the value of hyperparameter \( B \) under the premise that the historical information is known; i.e., \( \Theta = x, d/\rho, d/\rho \) is determined, which is expressed as \( E[Y (d)] | \Theta = x, d/\rho, d/\rho \) by mathematical formula.

If the hyperparameter \( \theta \) still changes further in \( d/\rho \) experiments/tests, then the conditional mathematical expectation of \( Y (d) \), the corresponding random variable, can be expressed as follows:

\[
E[Y (d) | \lambda, \rho] = \frac{1}{\rho} \frac{1}{\rho} \exp(-\lambda \rho d/\rho).
\]  

(34)

Then, further, we need to make sure that in the new experiment/test sequence, how much probability that hyperparameter \( \theta \) will change. According to our hypothesis in the second section, all the experiment/test sequence for each group in the initial state can keep a change state, the purpose of which is to ensure the overall model can be used to measure is in complete probability space, that is to say, in \( d/\rho \)/tests, hyperparameter \( \theta \) keeps changing the state of the conditional probability, which can be expressed as follows:

\[
P(\text{change} | \Theta = 0, d/\rho, \lambda, \rho) = 1.
\]  

(35)

As for \( (0, d/\rho) \) experiment/test interval, the conditional probability of the hyperparameter \( \theta \) remaining changing state can be approximately expressed, in the case that historical
information $\Theta = x, d_x, d_T$ and parameters $\lambda$ and $p$ have been determined:

$$P(\text{change}|\Theta = x, d_x, d_T, \lambda, p) = \frac{(1 - p)\exp[-\lambda(d_x - d_T)]}{p + (1 - p)\exp[-\lambda(d_T - d_x)]}.$$  

(36)

Among them, the molecules represented in $d_x$ tests still keep change state, but that in the experiment/test interval of $(d_x, d_T]$ keep zero step change probability, and the denominator is the probability that nothing has changed from the beginning of the experiment/test to the end of the historical information, as the case is in assumption (4) in Section 3.1.

Here, we apply a small trick of conditional probability. For the clever treatment of “1,” $[(1 - p)\exp[-\lambda(d_T - d_x)]]$

$$E[Y(d_i)|\Theta = x, d_x, d_T, \lambda, p]$$

\[= \frac{(1 - p)^{\lambda^x}\exp(-\lambda d_T)(1/p - 1/p\exp(-\lambda d_i)/p)}{L(\lambda, p|\Theta = x, d_x, d_T)}\]

$$= \frac{p^{-1}(1 - p)^{\lambda^x}\exp(-\lambda d_T) - p^{-1}(1 - p)^{\lambda^x}\exp[-\lambda(d_T + p d_i)\exp(-\lambda d_i)]}{L(\lambda, p|\Theta = x, d_x, d_T)}$$

(38)

It should be noted that the case of A1 when $x = 0$ is eliminated, because the premise for the existence of the entire posterior probability and the posterior mathematical expectation is that the entire historical information should remain in the new interval $(d_T, d_T + x]$ state at $d_T$, and the hyperparameter $\theta$ must be able to connect with each other.

Because the above results are involved parameters $\lambda$ and $p$, they are still incomplete expressions of random variable $Y(d_i)$ on posterior mathematical expectation. Therefore, according to Section 2 of model assumptions, $g(\lambda|r, a) = (a^\lambda r^{-1}\exp(-\lambda a)/L(r)), \lambda > 0$, and $g(p|a, b) = (p^{a-1}(1 - p)^{b-1}/B(a, b)), 0 \leq p \leq 1$; further processing, since the probability distribution density of these two parameters is continuous, we can adopt the method of integral and further improve A3 posterior mathematical expectation and add more uncertainty information, especially the hidden information about parameters $\lambda$, $p$ that express explicit use of parameter set $(r, a, b)$:

$$E[Y(d_i)|\Theta = x, d_x, d_T, r, a, b] = \int_{0}^{\infty} E[Y(d_i)|\Theta]$$

$$= x, d_x, d_T, \lambda, p \cdot g(\lambda, p|\Theta = x, d_x, d_T, r, a, b)d\lambda dp.$$  

(39)

where $g(\lambda, p|\Theta = x, d_x, d_T, r, a, b)$ is the joint posterior probability distribution density of the parameters $\lambda, p$.

According to the basic theory of Bayesian estimation and Bayesian optimization, under the condition that the parameters $\lambda$, $p$ is independent, their joint posterior probability distribution density can be expressed in the following complex form:

$$g(\lambda, p|\Theta = x, d_x, d_T, r, a, b)$$

\[= \frac{L(\lambda, p|\Theta = x, d_x, d_T)g(\lambda|r, a)g(p|a, b)}{L(r, a, b|\Theta = x, d_x, d_T)}\]

(40)

By combining the key intermediates variables A5 and A3 in formula A4, the posterior mathematical expectation related to the hyperparameter $\theta$ can be further deduced:

$$E[Y(d_i)|\Theta = x, d_x, d_T, r, a, b]$$

\[= \frac{A - B}{L(r, a, b|\Theta = x, d_x, d_T)},\]

(41)

where two parameters $A$ and $B$, respectively, represent the following two operation results:
For $A$,  
\[ A = \int_0^1 \int_0^\infty \alpha^{-1} (1 - p)^{\alpha - 1} \exp(\lambda_2 (e_1 + \lambda) g(\lambda_1, \alpha) g(p) a, b) d\lambda dp \]

\[ = \frac{B(a - 1, b + x)}{B(a, b)} \frac{\Gamma (r + x) a^x}{\Gamma (r) (a + d_r + d_i)^{r+x}} \]

(42)

For $B$,  
\[ B = \int_0^1 \int_0^\infty \alpha^{-1} (1 - p)^{\alpha - 1} \exp[-(\lambda_2 (e_1 + \lambda) g(\lambda_1, \alpha) g(p) a, b) d\lambda dp] \]

\[ = \frac{1}{\Gamma (r) B(a, b)} \int_0^\infty \left\{ \int_0^\alpha \lambda^{-x-1} \exp[-(\lambda_2 (e_1 + \lambda) g(\lambda_1, \alpha) g(p) a, b) d\lambda] \right\} dp \]

\[ = \frac{\Gamma (r + x) a^x}{\Gamma (r) B(a, b)} \int_0^1 \alpha^{-2} (1 - p)^{b+x-1} (a + d_r + d_i)^{-r-x} dp. \]

(43)

To achieve formal matching with the Gauss hypergeometric function in the form of Euler integral, we need to further realize variable substitution, and let $q = 1 - p$ and also make corresponding change $dp = dq$ for the differential variable. Further calculation results are as follows:  
\[ \Gamma (r + x) a^x \]

\[ \Gamma (r) B(a, b) (a + d_r + d_i)^{r+x} \]

\[ \int_0^1 (1 - q)^{a-2} (q)^{b+x-1} \left( 1 - \frac{d_i q}{a + d_r + d_i} \right)^{-(r+x)} dq. \]

By combining the Gauss hypergeometric function in the form of Euler integral, we get the parameter $r, b; a + b - 1; (d_i/a + d_i)$, which corresponds to the corresponding parameter expression in the form of integral function respectively, and we get:  
\[ \frac{B(a - 1, b + x)}{B(a, b)} \frac{\Gamma (r + x) a^x}{\Gamma (r) (a + d_r + d_i)^{r+x}} \]

\[ \frac{F_i (r + x, b + x; a + b + x - 1; d_i/a + d_r + d_i)}{\Gamma (r) B(a, b)} \]

(45)

Under the condition of formula A6, according to the mathematical expression formula of $L(r, a, b; \Theta = x, d_r, d_i)$ (6), the middle of the two key variables $A7$ and $A8$, after reduction, we can get the ideal result, under the BP/NBD model; the mathematical expectation of overall $\theta$ hyperparameters under the posteriori probability has analytical solution, and the mathematical expression of the specific is as follows:

\[ E[Y (d_i)|\Theta = x, d_r, d_i, r, a, b] \]

\[ = \frac{(a + b + x - 1/a - 1) [1 - ((a + d_r)/(a + d_r + d_i))^r] F_i (r + x, b + x; a + b + x - 1; (d_i/a + d_r + d_i))}{1 + I_{x>0} (a/(b + x - 1))(a + d_r/a + d_r)^{r+x}} \]

(46)

4. Experimental Analysis

4.1. Dataset of an Industry Competition. The premise of text mining of the entries requires complete project information:

(1) Most of the projects in the prototype design stage are in the initial stage, and there are some problems such as incomplete application materials. In this mining modeling process, the projects in the prototype design stage are eliminated.

(2) Projects without declaration materials (project documents) cannot pass the preliminary examination. In this mining modeling process, enterprises without declaration materials will be eliminated.

Finally, 143 project contents are reserved for topic mining, and the case content and project output value are mined separately to extract important information and explore the application field, current situation, and key points of value promotion of industrial Internet.

The main purpose of using this dataset is to adjust the discrete hyperparameters of the relevant prediction or classification model and compare the differences between the results before and after the adjustment, to form the experimental conclusion.

4.2. Participle. Jieba Chinese word segmentation component [10–12] is called, industrial Internet-related special words such as VR, AR, and smart Park are added into Jieba library, and precise mode is used to segment each document by default. In the process of word segmentation, function words and meaningless symbols are removed. To unify the expression of professional vocabulary, a thesaurus is set to merge synonyms to improve the effect of later topic extraction. For example, “artificial intelligence” and “AI” are synonyms and merged into “Ai.” The unimportant words are
filtered according to the part of speech, and the related words are finally reserved.

4.3. LDA Topic Extraction. The generated dictionary and corpus are in accordance with the input format of the model, and the open-source Gensim package [13] is used to construct the topic model and estimate the parameters. The default value of parameter selection is $a = 0.37$ and $B = 0.02$. The optimal topic number $k$ is determined according to the perplexity of the model. In theory, the $k$-nearest neighbor with low confusion degree should be selected, but under the premise of small corpus database capacity, more topics may lead to overfitting phenomenon. Therefore, this time, the number of topics is adjusted by combining the coincidence degree of the visualization results. Finally, the optimal number of topics is 4, the optimal number of output value topics is 2, and there is no coincidence between topics.

4.4. Basic Model. According to the results of index analysis and LDA theme mining [14–16], the projects that are in the prototype design stage and lack of application materials are selected to be eliminated. Finally, 143 projects are retained to enter the modeling stage, of which 67.83% can enter the second round. Considering that there are many projects in the preliminary evaluation stage, which are not easily affected by external factors and can better reflect the development trend of current related industries, the paper uses XGBoost and multinomial NB to predict and model the preliminary evaluation results of 2020 related industry competition. The discrete hyperparameter indexes of the correlation model are the highest polynomial degree (multinomial NB) and the number of root nodes (XGBoost).

4.5. Evaluation Index. Because of the classification model prediction, the confusion matrix is selected as the basic evaluation index, and the evaluation index is derived from confusion matrix, as given in Table 1.

4.6. Comparison of Experimental Results. In this modeling process, the polynomial naive Bayesian model (multinomial NB) and integrated model (XGBoost) are selected as training models (lack of specific mathematical description of the two methods). 70% of the samples are used for model training and 30% for testing. The parameters are adjusted in the training set dataset combined with threefold cross-validation, and “AUC” is used for training in the parameter adjustment process. Figures 1 and 2 show the effect of the model after adjusting the parameters, which has a certain reference value for the actual prediction performance of the model. On the test set, the effect of the model is as follows.

Through the comparison of experiments, we can find that

(1) From the classification prediction results of the whole model, the final result of XGBoost method is obviously better than the traditional Bayesian model in both the original configuration and the adjusted configuration

(2) The method of adjusting discrete hyperparameters based on skew distribution is effective. Through the empirical results of data experiments, the overall model, whether XGBoost or Bayesian model, achieves the improvement of relevant evaluation indicators under the new hyperparameter adjustment strategy.

<table>
<thead>
<tr>
<th>Classification</th>
<th>Forecast positive class</th>
<th>Forecast negative class</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actual positive class</td>
<td>TP—true positive</td>
<td>FN—false negative</td>
</tr>
<tr>
<td>Actual negative class</td>
<td>FP—false positive</td>
<td>TN—true negative</td>
</tr>
</tbody>
</table>

Table 1: Confusion matrix.

(1) Accuracy (ACC): $\text{Accuracy} = (TP + TN)/(TP + FN + FP + TN)$, the percentage of the correct results. (2) Precision: $\text{Precision} = TP/(TP + FP)$, In the records with positive prediction, how many are actually positive. (3) Recall rate (recall): $\text{Recall} = TP/(TP + FN)$; in the actual positive records, how many predictions are positive. (4) F1 score: $2/(1/\text{Precision} + 1/\text{Recall})$. 

![Figure 1: Results using original discrete hyperparameters.](image1)

![Figure 2: Results after adjusting the discrete hyperparameters using the expected results of skew distribution.](image2)
5. Conclusion

Through the combination of mathematical analysis and empirical research, this study focuses on the optimization method of the hyperparameters of the relevant machine learning model under the condition of skew distribution, especially the discrete hyperparameters. Through strict mathematical assumptions and mathematical derivation, the expected value of the discrete hyperparameters is obtained, and then, the actual data test is carried out through the empirical dataset. The experimental results show that this heuristic parameter adjustment method is feasible and effective for XGBoost and Bayesian models.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The author declares that there are no conflicts of interest.

References

[13] E. Mulholland, P. M. Kevitt, T. Lunney, J. Farren, and J. Wilson, “360-MAM-Affect: sentiment analysis with the google prediction API and EmoSenticNet,” in *Proceedings of