Research on roller coaster ranking model

Qianran Sun

International School of Beijing, China

Keywords: combination weighting method; TOPSIS.

Abstract: This paper establishes an objective evaluation algorithm for roller coaster ranking the data of roller coaster is first cleaned and normalized. Five evaluation indicators were selected by consumption level. The combination weighting method of combining the entropy method and the coefficient of variation method is used to determine the weight of the indicators. The combined weights of the indicators are obtained: W = (0.2873, 0.4702, 0.1866, and 0.056). A comprehensive evaluation model of the improved rank-sum ratio method and the TOPSIS method based on Mahalanobis distance was established, and the roller coaster was comprehensively evaluated.

1. Introduction

1.1 Background

With the improvements of the people's living standard and the stimulation and physical fitness function riding a roller coaster itself has, riding a roller coaster has become a sport loved by the majority of the people.

When people go to the park, riding a roller coaster is getting more and more popular. For those who are interested in roller coasters but have not touched a roller coaster or rarely touched a roller coaster, an objective and reasonable Roller Coaster rating/ranking site is very helpful for them.

Several online roller coaster rating/ranking sites rely heavily on subjective input to determine the rating or ranking of a particular roller coaster while considering some objective metrics (for example, a ”specialist” driver's ”thrill” or ”experience” score to measure ”thrill”).We need to address this problem.

1.2 Outwork

There are several roller coaster ranking websites on the Internet. They are not objective enough to evaluate the roller coaster ranking. It asks us to solve this problem. Before building the model, we first cleaned and normalized the data. Then, in order to absorb the advantages of different models and reduce the error, we combined the entropy method with the coefficient of variation method, used the combination weighting method to determine the weight, then the improved rank-sum ratio evaluation model and the TOPSIS comprehensive evaluation model based on Mahalanobis distance are used to comprehensively evaluate the roller coaster.

2. Assumptions

The following basic assumptions are made in order to simplify the problems and clear some vague concepts.

•The data we have collected is accurate and reliable.

•The data provided in the problem is accurate and reliable.

•Suppose we do market research covering all the roller coasters that exist in the world.

•It is assumed that the choice of all indicators is objective and has nothing to do with personal subjective factors.

3. A set of objective quantitative algorithms for roller coaster ranking

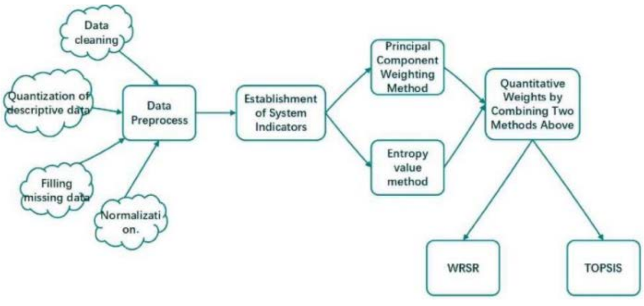

Figure 1: The overall chart which shows the ideas about our model

3.1 Model Preparation: Data Preprocess

Step 1: Data cleaning.

After analysis, it is found that some of these data are impractical, which will cause dis- tress to the later thinking and analysis. So it is necessary to delete or improve this part of the wrong data properly.

Table 2: Fault data

Column	Row	Reason	correction
H	57	Data and meaning are not right	Delete
H	64	Data and meaning are not right	Delete
H	265	Data and meaning are not right	Delete

Step 2: Quantization of descriptive data.

A large part of data given in the problem is descriptive data, which is now present¬ed in the form of text and cannot be quantitatively analyzed. In order to prepare for the subsequent quantitative analysis, it is necessary to quantify the text data that will be used later in this article. For example, geographical location, materials and so on.

•Quantization of geographical location

Different geographical locations represent different levels of consumption. People's spending budgets will affect their intuitive experience of roller coasters, so they need to be numerically located. The GDP of a region can represents the level of consumption there, so GDP can be used as numerical geographic location data to replace the raw location information.

•Quantization of materials

It can be found that the material only has two kind: steel and wood. After searching for data, it can be known that steel is safer than wood. If the sum of steel roller coaster safety factor and wood roller coaster safety factor is set to 1, then the steel roller coaster safety factor can be weight as 0.6, wood roller coaster safety factor can be weight as 0.4.

Step 3: Filling missing data.

It's easy to see that the data is missing part of it. The lack of data can have an impact on the subsequent study. Through careful analysis, it can be concluded that for missing data, filling the mean value is reasonable. This kind of data processing is helpful to increase the rationality and validity of later research.

Step 4: Normalization.

The data units and sizes given in the problem are very different, which will cause distress to the analysis of the latter text. So the numerical data and the quantized descriptive data should be normalized, which can be convenient for the study of the latter text.

3.2 Establishment of System Indicators

There are too many comprehensive evaluation indicators for roller coaster. More¬over, they often need more data which may hard to collect. According to the comprehensive and accurate principle, combined with the selection design scheme of the first problem, there are five indicators are proposed, which are shown in the figure. Introduction of each indicator.Risk Coefficient is negative data, others are forward data.

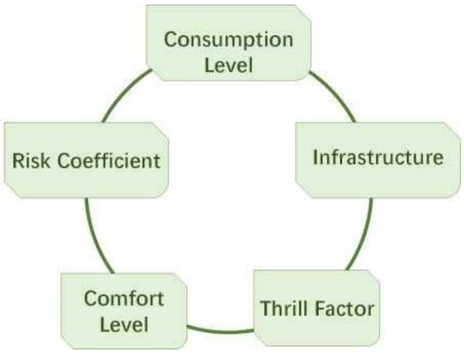

Figure 2: Five evaluation indexes of the Roller Coaster

•Consumption Level

Different groups of people have different levels of consumption. When people think about taking a roller coaster ride, they need to estimate of how much will it takes to ride a roller coaster. Roller coasters built in different regions, the cost of riding is not the same, even a large gap, which is very much related to the

local level of consumption. Therefore, the consumption level is selected as an indicator to objectively evaluate the roller coaster, and the amount of the indicator is related to the local GDP per capita and the Gini coefficient. This indicator can be calculated as follows:

$$ZX = GDP * Gn \quad (1)$$

Where ZX represents the value of the Consumption Level, GDP represents the value of the local GDP per capita, Gn represents the value of the Gini coefficient.

•Risk Coefficient

The roller coaster is a thrilling amusement facility. There will be a lot of dangerous movements during the ride, so ensuring the safety of passengers is a factor that needs to be considered first. Therefore, the risk coefficient is selected as an indicator to objectively evaluate the roller coaster. After analysis, it can be found that the sooner the roller coaster opens, the older the device, that is, the more dangerous. At the same time, the stronger the material, the safer it is. There- fore, the opening year and the material of the roller coaster are important factors that determine the risk coefficient. By looking up the data, we have determined the parameters. This indicator can be calculated as follows:

$$ZR = 1.37yec \quad (2)$$

where ZR represents the value of the Risk Coefficient, y represents the value of the opening year, c represents the value of the material of the roller coaster.

•Comfort Level

People enjoy the pleasure of riding a roller coaster, the comfort of the roller coast¬er will greatly affect people’s mood. It is only when the roller coaster is highly comfortable that people can play more happily. Therefore, the comfort level is chosen as an indicator to objectively evaluate the roller coaster. It can be found that the length of the roller coaster ride has the greatest impact on the comfort of people on the roller coaster ride. By looking up the data, we have determined the parameters. This indicator can be calculated as follows:

$$ZC = l + 0.2du \quad (3)$$

where ZC represents the value of the Comfort Level, l represents the value of the length of the roller coaster ride, du represents the value of the duration of the roller coaster ride,

•Thrill Factor

People who like or want to ride a roller coaster are hoping for a thrilling jour¬ney. If a roller coaster doesn’t give people the thrill of adventure, the existence of this roller coaster will be very low. Therefore, thrill factor is chosen as an ob¬jective evaluation of roller coaster indicator. It can be found that the Speed, G Force, Ver- tical Angle, Height and Number of Inversions of the roller coaster are proportional to the thrill factor. This indicator can be calculated as follows:

$$ZT = sp \times gf \times va \times he \times iv \tag{4}$$

where ZT represents the value of the Thrill Factor, sp represents the value of the Speed of the roller coaster, gf represents the value of the G Force of the roller coaster, va represents the value of the Vertical Angle of the roller coaster, he rep¬resents the value of the Height of the roller coaster, iv represents the value of Number of Inversions of the roller coaster.

•Infrastructure

The infrastructure of the roller coaster can affect the intuitive feeling of passengers. Therefore, the selection of infrastructure as an objective evaluation of roller coaster indicator. The factors influencing the infrastructure indicator of the roller coaster include Height, Length and Speed of the roller coaster. This indicator can be calculated as follows:

$$ZI = le \times sp \times he \tag{5}$$

where ZI is the value of the Infrastructure, sp is the value of the Speed of the roller coaster, le represents the value of the Length of the roller coaster, he represents the value of the Height of the roller coaster.

3.3 Weight Determination of Indicators

3.3.1 Quantitative weights by Entropy value method

Entropy is an indicator to measure the disorder degree of a system in information field, which can measure the effective information from data.Entropy method is a method to determine the weight of each index according to the information transferred to the maker of decision .The larger the difference between evaluation indexes is, the smaller the entropy is, the more information the indexes contain and transmit, and the larger the corresponding weight is.The specific steps showing follows:

Step 1: normalize the values of each indicator:

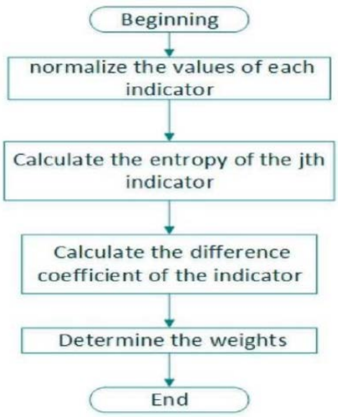

Figure 3: Calculation Steps of Entropy value method

$$aij = xij/\sum_{i=1}^{n} xij,\ i = 1, 2, 3, \cdots n,\ j = 1, 2, 3, \cdots m \tag{6}$$

Step 2: Calculate the entropy of the j th indicator:

$$Hj = -k\sum_{i=1}^{n} aij \ln aij(k = 1/1nn),\ k > 0,\ Hj > 0 \tag{7}$$

If xij is equal for all given j, then $aij = \frac{1}{9}$, at this time Hj = kln 9

Step 3: Calculate the difference coefficient of the indicator:

For given j, the less difference of xij the more Hj, when all of xij are equal, Hj = Hmax = 1(k = 1/1n), at this point for comparison between systems indicator xij has no effect; when the difference of xij is larger, Hj is smaller, the greater the effect of indicators on the system. Therefore, the difference coefficient is defined as:

$$gj = 1 - Hj \tag{8}$$

The bigger gj is,the more attention should be paid to the role of this indicator.

Step 4: Determine the weights:

$$w_j = \frac{g_j}{\sum_{i=1}^{9} g_i}, \ (j = 1, 2, \cdots, 4) \tag{9}$$

which is normalized weight coefficient. Thereby getting:

Table 3: Weight of each indicator

Indicator	ZX	ZR	ZT	ZT	ZI
weight	0.124	0.138	0.205	0.42	0.113

3.3.2 Quantitative weights by Coefficient of Variation Method

If all objects can be evaluated by an index value of an actual distinction,the index may get a larger weight value, otherwise, it may have a smaller one. This way is strongly reliable.

Suppose there are n evaluation objects, each object will be described with above indexes, which are expressed as xj1, xj2, · · · xjP(i = 1, 2, · · · , 6). Mean value Xi and variance Si2 should be calculated with below formulas.

$$\text{overline Xi} = \frac{1}{n}\sum_{j=1}^{n} xji \tag{10}$$

$$Si2 = \frac{1}{n-1}\sum_{j=1}^{n}\left(xji - xi\right)^2 \tag{11}$$

Then the coefficient of variation index can be computed.

$$V\ i = Si\sqrt{xi}\ (i = 1, 2, \cdots, P \tag{12}$$

Normalize V i and set ωi as objective weight of each index. It can be computed with below equation.

$$\text{omegai } 1 = \frac{Vi}{\sum_{i=1}^{5} Vi} \tag{13}$$

Then the objective weight of each index is obtained. The result is as follows:

Table 4: Weight of each indicator

Indicator	ZX	ZR	ZT	ZT	ZI
weight	0.163	0.105	0.147	0.37	0.213

3.3.3 Combination Weighting Method

The weights determined by Coefficient of Variation Method and Entropy value method reflect the importance of each indicator in different aspects. By combining the two methods, it is possible benefitting from both advantages.

$$W\ i = \frac{wi^{(1)} wi^{(2)}}{\sum_{j=1}^{p} wj^{(1)} wj^{(2)}}\ i = 1,2,\ldots 5 \tag{14}$$

The results of finalizing the indicator weights are as follows:

Table 5: Weight of each indicator

Indicator	ZX	ZR	ZT	ZT	ZI
weight	0.1435	0.1215	0.176	0.396	0.163

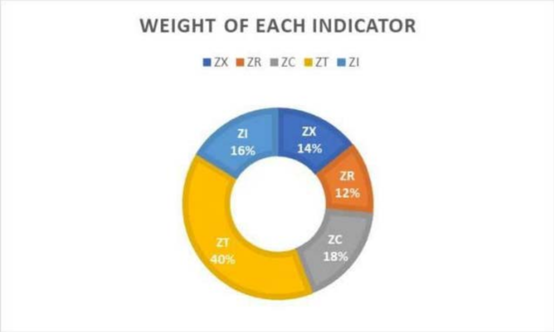

Figure 4: Weight of each indicator

Through objective analysis, among the five indicators, the weight of the stimulating factor is the largest. It can be known that the stimulating feeling brought by the roller coaster is the most important when people ride a roller coaster.

3.4 Quantitative Evaluation Algorithm

3.4.1 Improving the Roller Coaster Evaluation Model Based on Weighted Ranksum- Ratio Method (WRSR)

This model adopts the WRSR method. By correcting the indicator weight coefficients, the indicators are more accurately analyzed, and the dimensionless statistics are obtained by rank conversion. Using the method of parametric statistical analysis, the value of the evaluation object is comprehensively evaluated. However, in the process, information on the original indicators is inevitably lost, so improvements are made on this basis. The specific steps are as follows:

Step 1: Calculate Empirical Weight Coefficient

We use the Delphi method to establish empirical weight coefficients through experts.

Step 2: Assign Rank

Arranging m evaluation indicators of n evaluation objects into a raw data table of n rows and m columns. Assign the rank of j evaluation objects peri indicators. The resulting rank matrix is recorded as R∗ = (Rij)mn

$$R^* = \begin{bmatrix} R_{11} \cdots R_{1n} \\ \vdots \ddots \vdots \\ R_{m1} \cdots R_{mn} \end{bmatrix} \tag{15}$$

Step 3: Calculate the WRSR

When the weights of each evaluation indicator are different, suppose the weight of the j th indicator is ωj. And the W RSR is calculated as follows:

$$\text{WRSRi} = \frac{1}{n}\sum_{j=1}^{m} wjRij, i = 1,2,\cdots,n \tag{16}$$

Step 4: Calculate the Probability Unit

Arrange the W RSR values in order of small to large, list the frequency fi of each group, calculate the cumulative times cfi, and the cumulative frequency pi .

pi = cfi/n (17)

Convert pi to probability unit Probit, Probit is the standard normal distribution of pi’s quantile plus 5.

Step 5: Calculating the Regression Equation

Taking the probability unit Probit corresponding to the cumulative frequency as the independent variable, and W RSRi as the dependent variable, calculate the regression equation:

$W\ RSR_i = a + b \times Probit_I$ (18)

Step 6: Sort to Get Evaluation Results

According to the W RSR estimation value corresponding to the regression equation estimation, the evaluation objects are sorted, so that the roller coaster ranking is inferred by the sorting.

Step 7: Result Analysis

We use Matlab to solve the frequency fi of the roller coaster given in the data, the cumulative frequency pi, and the probability unit Probiti the results are shown in the table:

Table 6: Cumulative frequency, probability unit and estimated valueof WRSR

Roller Coasters	fi	cfi	pi	Probiti	W RSRfit	rank
Tower of Terror II	1	5	0.9643	6.8027	0.5595	4
Big Apple Coaster	1	4	0.8571	6.0676	0.5208	25
Extreme Rusher	1	3	0.7143	5.5659	0.4911	98
Top Thrill Dragster	1	2	0.5714	5.18	0.4633	151
T Express	1	1	0.4286	4.82	0.4336	193

Find the linear regression equation as:

$$WRSRi = 0.8626 - 0.0771 \times Proboti \tag{19}$$

The sorting results of each roller coaster are shown in the following figure:

3.4.2 TOPSIS Comprehensive Evaluation Model Based on Mahalanobis Distance

According to the analysis, there is a certain degree of correlation between the indicators, and the traditional TOPSIS method can not effectively deal with the common information between the indicators. Therefore, we establish a TOPSIS method based on the Mahalanobis distance evaluation model. The variance matrix responds to the correlation of different decision indicators and can effectively eliminate the duplication information.

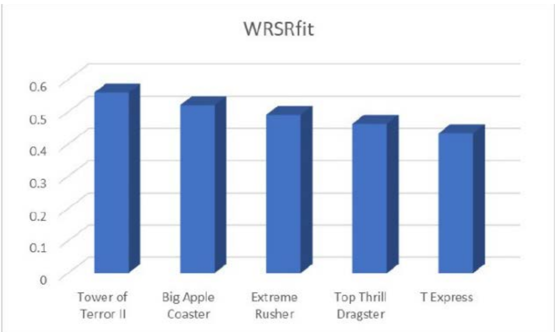

Figure 5: The RSRfit value of the top seven roller coasters

Step 1

First, establish the decision matrix for this problem according to the general steps of the TOPSIS method. The decision matrix is denoted as A, which can form the standardized decision matrix Z, whose elements are denoted as Z1ij, and have where fij is given by A, which means the judging value for the j th indicator for the ith object.

$$Z1ij = \frac{f_{ij}}{\sqrt{\sum_{i=1}^{n} f_{ij}^2}}, i = 1,2,\cdots,5,\ j = 1,2,\cdots,6 \tag{20}$$

Step 2

This paper randomly selects some normalized decision matrices. According to the survey, it is found that people care about each indicator the same, so their weights are equal, so the weighted matrix is approximated as:

$$Z = Z1 = \begin{pmatrix} 0.3912 & 0.0088 & 0.4512 & 0.2878 & 0.3454 \\ 0.1416 & 0.0456 & 0.3991 & 0.1268 & 0.4564 \\ 0.9606 & 0.0450 & 0.1518 & 0.2992 & 0.4001 \\ 0.4589 & 0.0921 & 0.0411 & 0.2783 & 0.1484 \\ 0.8958 & 0.0153 & 0.0614 & 0.2068 & 0.4255 \end{pmatrix} \quad (21)$$

Step 3

Because in the scheme, the larger the indicator value, the better, so the ideal solution and the negative ideal solution are:

$$\begin{pmatrix} 0.2452 & 0.0409 & 0.3451 & 0.2173 & 0.2635 \\ 0.5279 & 0.0456 & 0.0784 & 0.4026 & 0.1214 \end{pmatrix} \quad (22)$$

Distance measurement and solution of relative proximity. The distance from any feasible solution to Z+ is:

$$S_i^+ = \sqrt{\sum_{i=1}^{5}(Z_{ij} - Z_j^+)^2} \quad (23)$$

The distance from any feasible solution to Z− is:

$$S_i^- = \sqrt{\sum_{i=1}^{5}(Z_{ij} - Z_j^-)^2} \quad (24)$$

The relative proximity of an arbitrary feasible solution to an ideal solution is defined as:

$$C_i^* = \frac{S_i^-}{S_i^- + S_i^+} \quad (25)$$

The bigger the value, the better.

Table 7: Relative proximity of the indicators of different roller coasters to the optimal solution

Roller Coasters	S+	S-	C*	rank
Kingda Ka	0.1113	0.9173	0.8918	4
Extreme Rusher	0.227	0.7832	0.7753	52
Nemisis Inferno	0.2628	0.1162	0.3066	167
Adrenaline Peak	0.357	0.0796	0.1823	251
Wild Thing	0.5489	0.882	0.6164	123

Step 4 finally, each calculated Mahalanobis coaster over the positive and negative solutions over the distance Solution:

$$d_i^+ = (0.1508\ 0.5090\ 0.2826\ 0.3324\ 0.2431) \quad (26)$$

$$d_i^- = (0.5927\ 0.6432\ 0.1326\ 0.3256\ 0.7263) \quad (27)$$

4. Conclusion

This study employed the evaluation model mentioned above to rank the roller coaster and obtained reasonable results.

References

[1] Fang Wang, Wenfeng Shen. The review of design planning of Tourism Resort [J]. Foreign investment in China, 2009, (11).

[2] Wang Tao, Rong Lei. The Improved Algorithm for the Extraction of Contour Lines from GridDEM[J]. GeomaticsWorld, 2006, 4(1).

[3] Jiuping Xu, Jun Li. Multiple Objective Decision Making Theory and Methods [M]. Tsinghua University Press, 2005.

[4] Linear weighted sum method. https://en.wikipedia.org/wiki/linear weighted sum method.

[5] Coefficient of Variation Method. https://en.wikipedia.org/wiki/Coefficient of Variation method.

[6] Multi-correlation coefficient. https://en.wikipedia.org/wiki/Coefficient of Varia- tion Method.