Horse Racing Decision and Analysis Based on Data Mining and Big Data Analysis

Shuang Zhang[1] and Mairu Liu[2(✉)]

[1] Wuhan Business University, Wuhan, Hubei, China
[2] Central China Normal University, Wuhan, Hubei, China
mairu73info@eiwhy.com

Abstract. Nowadays, the horse racing betting market is developing rapidly, which makes the prediction of the outcome of horse racing has become a new challenge. This research mainly discusses horse racing decision-making based on data mining and big data analysis technology. The main functions of the system designed in this study can be divided into two main parts: publishing competitions and participating in competitions. Publishing the competition means that the organizer of the competition publishes its own competition through the system. The content of the release mainly includes the description of the competition, the time of the holding, the rewards, etc., and the most important part is that the organizer must provide the information used in the competition data; and participating in the competition means that the participants of the competition obtain all relevant information of the competition through this platform, the most important thing is the data of the competition, and after the data is processed by data mining, the prediction results of the data are obtained., And submit the results to the system, and the system will automatically score the results. For the first 10 games, the accuracy of predicting the winning percentage can reach up to 91%. This research will promote the development of the horse racing industry.

Keywords: Data mining · Big data · Horse racing decision-making · Outcome prediction

1 Introduction

Behind the bright appearance, the lack of cultural impetus for the development of the horse racing industry is an obvious fact. There are two reasons for the lack of this drive performance. One is the loss of horse culture in the folk culture, and the horse culture that affects the development of the city is almost exhausted. For the average person, the horse culture is just a book, and their remoteness from life and folk culture is bleak.

But now, with the rapid development of lottery business, many cities have built large-scale racetracks, and the software and hardware equipment environment is gradually improving [1, 2]. In order to adapt to the development of domestic sports companies, the expansion of the horse racing industry in this field promotes the development of physical culture and sports undertakings in a diversified direction, while also providing

new opportunities for sports science research, as well as for domestic The development of horseback riding has played a positive role [3, 4]. Related research in the horse racing industry is in its heyday. In recent years, data mining has been the main focus of research, and its development prospects may also be unlimited. However, apart from books and some authoritative websites that obtain necessary knowledge and external data from the Internet, there is almost no other way for general learners to obtain some latest data. Through the online platform, learners can download the latest data for free [5, 6]. For general data owners, if they are incapable or do not want to spend too much money, they can publish data on this platform. Participants can compete freely and analyze the results from excellent results. The prosperity of the domestic horse racing market has motivated the emergence of horse racing lottery tickets in our country [7, 8]. The basket lottery has achieved significant benefits in the United States and Europe where the horse racing market is developed. As an important part of sports betting, horse racing lottery not only injects new elements into the lottery market in our country, but also satisfies the desire of horse racing fans to participate in the game [9, 10].

Horse racing result prediction has become an important direction of current sports event result prediction research. The use of big data analysis methods can handle a large amount of data generated by the game well, and reduce the influence of subjective factors in artificial prediction on the prediction results. In this article, in addition to the basic features of home and away games, scores, etc., the prediction of the outcome of the horse racing competition also uses the basic information of the team participants in each game as the input features.

2 Horse Racing Decision

2.1 Horse Racing

So far, horse racing has made progress in business. In the era of people's spare time, spiritual and cultural life is developing day by day, and meeting people's increasing entertainment needs is a topic facing the world today. Gambling horse racing not only meets this demand of people, but also stimulates the economic source of entertainment horse racing and makes horse racing directly commercialized. In addition, the horse racing industry has also promoted the overall development of the horse racing industry and brought huge commercial benefits to economic development. National enterprises and ordinary people are working hard to invest in this new sports industry.

China's sports lottery market is closely integrated with sports games and is an essential element of the sports lottery market. With the progress of society, the improvement of people's quality, and the change of lifestyle, horse racing, as a kind of culture, has the characteristics of modern society. First of all, horse racing in many countries in the world is regulated by the government, and many are directly managed by the government. Japan directly takes the local government as its main body, and Singapore is also directly monopolized by the government. Although Hong Kong is organized and operated, it is legally recognized by the government. Second, it has a welfare function. Today, horse racing is a welfare undertaking, and the modern jockey club is a welfare organization. From the perspective of the development of the Hong Kong Jockey Club, more than 1 billion yuan of benefits are used for charity every year. Finally, it is a public welfare

charity organization with many cultural characteristics such as modern horse racing, leisure, entertainment, and history. It is closely related to transportation, catering, construction, etc., and is also a comprehensive social and cultural activity, which embodies the inclusion of people regional economic development including living standards and cultural needs.

2.2 Data Mining

When predicting the results of horse racing, some representative teams will randomly select and classify, and select their attributes as decision attributes by constructing decision trees, pruning decision trees, and extracting decision trees. The deletion of attributes, the transformation of attributes and the generalization of the continuous attribute data of the interval value, and the reasonable prediction of the outcome of the game according to the rules. The decision tree C4.5 algorithm is used to build a horse racing prediction model. This study designed a data set containing two attribute values. These are the scores of the last 5 rounds and the overall goal rate of the team. Then divide the results into three categories: wins, draws, and negatives. The decision tree is generated based on this data set of the C4.5 algorithm.

The input value S of the data mining process is:

$$S = \sum_{i=1}^{n} W_i X_i \tag{1}$$

The output value y is:

$$y = f\left(\sum_{i=1}^{n} W_i X_i - \beta\right) \tag{2}$$

Linear function conversion:

$$y = (x + \text{MinValue})/(\text{MaxValue - MinValue}) \tag{3}$$

Asynchronous means that the prediction result file uploaded by the user is not processed immediately, but stored in the message queue and processed by a process that starts in the background. Complete control of the process, including the master process and the slave process group. The master device controls the operation of the slave device. Asynchronous processing can not only appropriately control the server's CPU and memory usage, improve the stability of the system, but also improve the response speed and performance of the system. Modern mathematics, statistics, data mining and other fields are closely integrated with sports theory, in-depth technical and tactical analysis of sports competitions, forming a series of theoretical systems with various fields as the main body, providing scientific basis for decision-making. Selection of technical and tactical analysis methods for sports competitions. However, it is not enough to analyze technology and tactics from a single aspect of a certain field. In future research, in order to analyze sports technology and tactics, more scholars and experts are needed to conduct more detailed research.

3   Data Mining Experiment for Horse Racing Decision

3.1   Composition of the Horse Racing System

This article will introduce the design and implementation of an online data mining horse racing platform using PHP as the main development language. The system provides all the basic functions of a general online platform including user management systems such as new user application and user login. In order to facilitate communication between participants and the organizer, auxiliary functions such as forum management and broadcast management have been added.

3.2   Prediction of Match Results

At the same time, the design and implementation of the online horse racing system needs to have the function of publishing and submitting the competition, taking into account details and error handling. The game can be played within the prescribed range. In the evaluation of the submitted results, two different processing methods, real-time and asynchronous, are provided in the background. At the same time, after the user uploads the prediction result file to the server, the system will immediately evaluate and give a score.

4   Results and Discussion

The results of the age survey of horse racing are shown in Table 1. 28 people under the age of 20 accounted for 6.1%. 74 people from 20 to 30 years old accounted for 16.0%. 137 people aged 31–40 accounted for 29.7%. There are 106 people aged 41–50, accounting for 22.9%. There are 103 people aged 51–60, accounting for 22.3%. There are 14 people over 60 years old, accounting for 3.0%.

Table 1.   Results of the age survey of horse racing

Age	Number of people	Percentage	Cumulative percentage
Under 20	28	6.1	6.1
20–30 years old	74	16.0	22.1
31–40 years old	137	29.7	51.8
41–50 years old	106	22.9	74.7
51 to 60 years old	103	22.3	97.0
Over 60 years old	14	3.0	100.0

If horse racing is issued in China, the respondents believe that the price of each bet is 1 yuan, accounting for 15 people, accounting for 3.3%; the price of each bet is 2 yuan, accounting for 287, accounting for 62.1%; the price of each betting is 5 yuan, 152
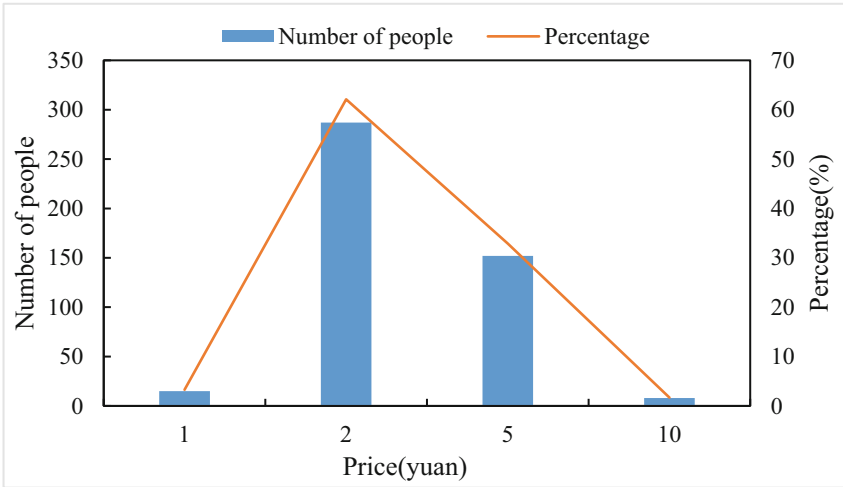
Fig. 1. The number of price identities per annotated horse race

people, accounting for 32.9%; 8 people with a price of 10 yuan per bet, accounting for 1.7%. Figure 1 shows the number of people who agree with the price of each horse race.

The prediction accuracy during the game is shown in Fig. 2. It can be seen that the prediction accuracy of the winning percentage in the first 10 games can reach up to 91%.
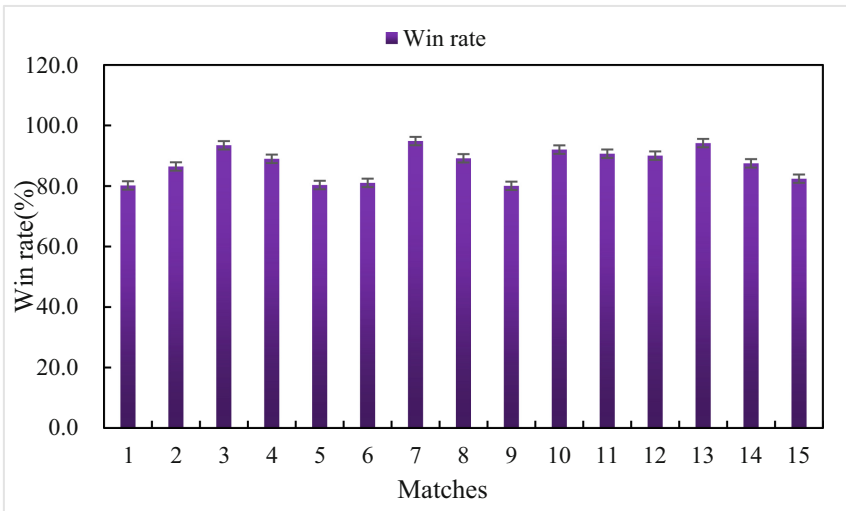

Fig. 2. The prediction accuracy during the competition

5   Conclusion

In the information age, the rapid development of computer technology, especially the high application of database technology, is one of the focuses of the rapid growth of large amounts of data, and how to effectively use this large amount of data to better serve human information science and technology. One of the main functions of data mining is to quickly and efficiently extract answers to questions from a large amount of actual data. Data mining uses a lot of basic data in finance, securities, commerce, insurance and other industries, but relatively few researches and applications in competitive sports technology and tactics. This research first analyzes the factors that affect the outcome of the game and their relationship with the outcome, and then tells the practicality of data mining for the horse racing prediction problem, and the ability to solve the problem of non-linear relationships. In the process of modeling the strength of the players, only the number of wins and losses is used as the basis for the construction of the voting matrix elements, and the other situations of the players, such as home and away, coaches, game strategies, etc., are not considered. Therefore, in the future can be discussed in depth in his work.

References

1. Lopez-Gonzalez, H., Griffiths, M.D.: "Cashing Out" in sports betting: Implications for problem gambling and regulation. Gaming Law Rev. **21**(4), 323–326 (2017)
2. Berry, C.E., Drummond, M.B.: The horse-racing effect and lung function: Can we slow the fastest horse? Am. J. Respir. Crit. Care Med. **195**(9), 1134–1135 (2017)
3. Cozier, J.G.: An economic analysis of the horse racing industry in Trinidad and Tobago. Soc. Econ. Stud. **67**(2/3), 177–214 (2018)
4. Mcswain, D.L.: The legalization of sports betting: A federalism framework and the horse racing model. J. Nat. Resour. Environ. Law **11**(1), 63–101 (2019)
5. Sacopulos, P.J.: Pitts vs barr-tonko bills: An in-depth comparison of proposed anti-doping legislation in horse racing. J. Nat. Resour. Environ. Law **9**(1), 37–64 (2017)
6. Nettlefold, J.: Horse racing takes a step into the future - excelitas qioptiq films with thermal cameras. Battle space C4ISTAR Technologies **20**(4), 30–33 (2017)
7. Han, H., Wen, Y., Chua, T.S., et al.: Toward scalable systems for big data analytics: A technology tutorial. IEEE Access **2**(2), 652–687 (2017)
8. Athey, S.: Beyond prediction: Using big data for policy problems. Science **355**(6324), 483–485 (2017). [Special Issue Perspective]
9. Kuang, L., Hao, F., Yang, L.T., et al.: A tensor-based approach for big data representation and dimensionality reduction. IEEE Trans. Emerg. Top. Comput. **2**(3), 280–291 (2017)
10. Janssen, M., Haiko, V., Wahyudi, A.: Factors influencing big data decision-making quality. J. Bus. Res. **70**, 338–345 (2017)