Research on the Intelligence of NBA2K Game Based on Machine Learning

Hongyu Lian

Fujian Normal University, Fuzhou, Fujian, China

Abstracts: With the rapid development of machine learning techniques, their application in various fields is becoming increasingly widespread. Data-based machine learning is an important aspect of modern intelligent technology, which studies the search for laws from observed data (samples) and uses these laws to make predictions about future data or unobservable data. The aim of this thesis is to explore the possibility of applying machine learning techniques to the NBA2K basketball game and to investigate their impact on game intelligence. By analysing game data and patterns, models are built using machine learning algorithms and integrated into the game system in order to improve the playability and challenge of the game and to achieve a more personalised gaming experience.

Keywords: Machine Learning; NBA2K; Data Analysis; Game Intelligence

1. Introduction

The main thrust of machine learning research is the use of computers to simulate human learning activities; it is the study of the ways in which computers recognise existing knowledge, acquire new knowledge, continually improve their performance and achieve their own refinement[1] Machine learning is an important branch in the field of artificial intelligence, which has made rapid progress in recent years. Machine learning is an important branch in the field of artificial intelligence and has made rapid development in recent years. Firstly, the algorithms and models of machine learning are constantly advancing. Traditional machine learning methods, such as decision trees and support vector machines, have achieved great success in application. In recent years, deep learning, on the other hand, has become a hotspot for machine learning, achieving excellent performance through multi-layer stacking of neural network structures and large-scale data training. This has enabled machines to make great breakthroughs in image recognition, speech recognition, natural language processing, etc., and has also promoted the application of related technologies in medical, financial, and transport fields.

Second, the abundance of data plays a key role in the development of machine learning. With the popularity of the Internet, the amount of data generated and stored by people has increased dramatically. These huge amounts of data provide rich training samples for machine learning and help machine learning algorithms to understand and predict better.

Machine learning has potential for game outcome prediction. By training models to analyse large amounts of historical data, machine learning algorithms can learn the correlation between different factors in a game and the outcome, and be able to predict future game outcomes.

For example, in board games, machine learning can learn patterns of victory or defeat by analysing past games and opponents' moves, and provide the best strategy for the next move. In eSports games, machine learning can predict the outcome of a match by analysing data such as a player's actions and team performance, and provide real-time advice to the player.

However, it is important to note that the outcome of a game is affected by a variety of factors, including player skill level, strategy, luck, etc. Machine learning is only a tool and its predictions may not be 100 per cent accurate. In addition, the randomness of the game is an important consideration, as machine learning algorithms are better suited to deal with deterministic patterns.

This post will focus on NBA2k 23 and analyse its predictions for the outcome of the game.

2. Introduction to the Game

NBA2k's Manager Mode compares to other games in that its game content and game orientation is completely analysed by the AI based on realistic player data and simulates the outcome of the game. The player can participate in the management of the team, such as the lineup, salary, etc., and the results of the game will be different according to a variety of factors.

Gameplay: In the game, players will play the role of a certain NBA team's general manager, running all the affairs of the team.

There are two main periods in which players operate: the start of the season through the trade deadline and the offseason

Start of season to trade deadline:

During this period, as the team is forming and playing frequently, the player is mainly responsible for talking to the players or the owner according to the prompts, as a way to increase the trust between the players and the owner (if the owner's trust is too low, the player will be dismissed), and to improve the team's chemistry in order to increase the probability of winning.

In addition, players can trade players with other teams, and the success of the trade is based on the player's ability value, salary and other aspects.

Offseason:

At this stage, the boss will first give you tasks to complete, which will increase your trust. Tasks include the search for good coaches, financial officers and other personnel, the signing of good free agents, the renewal of the contract of the players in the team and so on.

Player Ratings and Recruiting: In Manager Mode games, players need to recruit players by signing and procuring draft picks, and line up players based on their abilities. This process involves a lot of data and statistical analysis, in which machine learning techniques can be used to help calculate players' ability values and ratings, and predict their future performance. These predictions can inform managers to make more informed decisions.

Player trading: In manager mode games, players also need to make targeted adjustments to their team lineups based on season conditions, player status, and other factors. This adjustment involves player trading. Traditional player trading relies on human experience and perceptual judgement, but with the help of machine learning technology,

trading information and historical data can be better mined to make better decisions in an automated way.

Tactics and Match Management: In Manager Mode games, players also need to choose the right tactics and match strategies to get their team to win. Different tactics and strategies are affected by many factors, such as the rules of the game, players' abilities, venues, opponents, and so on. Machine learning techniques can be used to parse these complex data patterns and provide players with individually customised tactics and strategies.

3. Explore the Correlation Between High and Low Player Scores and Each Value

Here I will use the data within the NBA2k23 version as a reference to simulate the performance of the players in the playoffs in the game, and analyse the values of each of its games.

Within the game, I ran 15 simulations using a playoff format.The results of the 15 simulations are as follows.

Table 1. Results of the 15 Simulations

sports team (basketball, soccer, football etc)	Number of championships won
Boston Celtics.	7
Atlanta Hawks	6
New Orleans Pelicans	2

(Player injuries were turned off in all simulations to minimise simulation errors, as a way of ensuring the integrity of the team personnel)

From the simulation, it can be seen that the Boston Celtics and Atlanta Hawks have the highest probability of winning the championship, and the core players of their teams, Jayson Tatum and Trae Young, have been awarded the title of Most Valuable Member of the Finals, respectively. The game averages for both players are listed below.

Table 2. Game Averages for Both Players

	Jayson Tatum	Trae Young		Jayson Tatum	Trae Young
score	33.0	28.0	caps	1.1	0
offensive rebound (basketball)	3.8	0.4	fault	3.1	2.1
backboard	12.1	2.4	hit rate	50.6	44.1%
assist	4.6	10.9	Three-point field goal percentage	42.4 percent	40.9 percent
seize (an opportunity)	1.4	1.1	free throw percentage	90.3 percent	86.0 percent

(i) Traditional linear regression models

The traditional linear regression model can be used to predict outcomes for continuous variables. It is based on a linear relationship between the independent and dependent variables and makes predictions about new data by fitting an optimal straight line.

Linear regression models are usually used to predict the outcome of continuous variables, whereas the outcomes of NBA 2K games are usually discrete, such as the winners and losers of games or team scores. Therefore, it may not be appropriate to use linear regression models to simulate the outcomes of NBA 2K games.

However, linear regression models can be applied to predict how well a player will score in a game. Its predictors include, but are not limited to, shooting percentage, shot preference, foul making ability, number of shots taken, etc. A player's field goal percentage can be roughly calculated as point=2mx+3ny+b (m and n are two- and three-point shots, respectively, x and y are the number of shots taken, and b is the number of points scored on free throws)

The NBA's Senior Vice President Dean Oliver invented the EPM (Estimated Plus Minus) in 2004 as a way to measure a player's impact on the game.Dean Oliver is a basketball statistician and analytics expert who has held several positions in the NBA and has made significant contributions to the statistical analysis of basketball.The EPM of an NBA player is a statistical measure of a player's overall efficiency. Estimated Plus Minus (EPM) is a statistical measure of a player's overall efficiency.EPM is calculated by using a multiple linear regression model to analyse the change in a team's net scoring when a player is on the court, in order to assess the player's impact on a team's chances of winning or losing.

Specifically, the process of calculating the EPM value is as follows:

1. First, the team's Net Rating was analysed using a regression model, with Net Rating as the dependent variable and each of the player's presence statistics as the independent variable.

2. In a regression model, the independent variables can include statistical indicators such as a player's points, rebounds, assists, steals, caps, and turnovers.

3. The regression model calculates the coefficient of influence of each statistical indicator on the team's net score, i.e. the weight of each indicator.

4. The values of the statistical indicators when each player is present are substituted into the regression model and the predicted value of each player's contribution to the team's net score is calculated according to the weights.

5. Finally, the difference between the net points scored with each player present and the predicted value is calculated as the EPM value for that player.

It is important to note that the EPM value is only an estimate of a player's overall efficiency and is not an absolutely accurate measure. Since basketball is a teamwork game, a player's EPM value is affected by many factors, including team tactics, level of teammates, intensity of the game and so on. Therefore, EPM value can only be used as a reference, combined with other statistics and observations to evaluate a player's performance.

(ii) Decision trees

A decision tree is a tree-like structure that starts at the root node, tests data samples (consisting of sets of instances instances have a number of attributes), and divides the data samples into different subsets of data samples based on different results, each of which constitutes a child node[2] . It is feasible to build a decision tree model to predict the outcome of NBA2K game. Decision tree is a commonly used machine learning algorithm that learns from historical data to generate a tree-like structure that can be used to predict the classification or outcome of new data. To build a decision tree model to predict NBA2K game outcomes, the following steps will be required:

1. Data Collection: Collection of data related to the outcome of the NBA2K game. This data can include a team's historical win/loss record, players' skill attributes, and the history of matchups between teams.

2. Data preparation: Clean and organise the collected data. Ensure the completeness and accuracy of the data and transform it into a format suitable for processing by the decision tree model.

3. Feature selection: Select features that are relevant to the prediction of game outcomes. For example, the team's record, players' scoring ability, rebounding ability, etc. might be selected as features.

4. Data partitioning: dividing the data set into a training set and a test set. Typically, the training set is used to train the decision tree model and the test set is used to evaluate the performance of the model.

5. Model training: Use the training set data to construct a decision tree model. The decision tree algorithm will generate a decision tree structure based on the characteristics and results of the data.

6. Model evaluation: The performance of the model is evaluated using test set data. Some metrics such as accuracy, precision, recall, etc. can be used to evaluate the predictive power of the model.

7. Model optimization: Based on the results of the assessment, the model is optimized. Attempts can be made to adjust model parameters, select different features, or use other algorithms for comparison.

8. Predicting results: Finally, a trained and optimised decision tree model was used to predict the results of the new NBA2K game.

It is important to note that the decision tree model, as a method in machine learning, still has some limitations in predicting the outcome of NBA2K games [3]. The outcome of the game may be affected by many factors, including player status, tactical layout, climate, and so on. Therefore, relying only on historical data and decision tree models to make predictions may not be completely accurate [4].

4 Application and Limitations of Machine Learning Models in NBA2K Games

4.1 Traditional Linear Regression Models

The traditional linear regression model can be used to predict how well a player will score in a game. However, it also has some limitations. Firstly, the linear regression model assumes that there is a linear relationship between the independent and dependent variables, which may not always hold true in practice. Second, the linear regression model is sensitive to outliers, which may affect the model's prediction if there are large outliers in the data [5]. To improve the performance of the linear regression model, the following optimisation methods can be considered:

•Data pre-processing: Cleaning and organising the data, removing outliers and missing values to ensure the accuracy and completeness of the data.

•Feature selection: features that are highly correlated with the player's score are selected for modelling, such as shooting percentage, number of shots, etc.

•Model tuning: reduce the overfitting phenomenon of the model and improve the generalisation ability of the model by adjusting the parameters of the model, such as regularisation coefficients.

In addition, data enhancement methods can be used to increase the number of samples and improve the training effect of the model. For example, more training data can be generated by simulating different game scenarios and player states.

4.2 Decision Tree Modelling

Decision tree modelling is a commonly used machine learning algorithm that learns from historical data to produce a tree-like structure that can be used to predict the classification or outcome of new data. However, decision tree models have some limitations. For example, decision tree models are prone to overfitting, especially when there is less training data or more features. In addition, the prediction results of decision tree models may be affected by feature selection and data preprocessing.

To improve the performance of the decision tree model, the following optimisation methods can be considered:

•Feature selection: Select features that are highly relevant to the game outcome to be modelled, such as team record, player skill attributes, etc.

•Pruning strategy: pruning strategy to reduce the complexity of the decision tree, to prevent the occurrence of overfitting phenomenon. Commonly used pruning strategies include pre-pruning and post pruning.

•Integration learning: integrating multiple decision tree models, such as Random Forest and Gradient Boosting Tree, to improve the predictive power and stability of the model.

In terms of data pre-processing, the collected data need to be cleaned and organised to ensure the completeness and accuracy of the data. For example, missing values can be filled or deleted, outliers can be handled, and so on. In addition, operations such as encoding and scaling of features are required to meet the requirements of the decision tree model.

4.3 Assessment of Indicators

In order to comprehensively assess the prediction performance of the model, evaluation metrics such as accuracy, recall and F1 score need to be introduced. Accuracy is the ratio of the number of samples correctly predicted by the model to the total number of samples; recall is the ratio of the number of actual positive samples predicted by the model to the number of all actual positive samples; and F1 score is the reconciled average of accuracy and recall, which is used to comprehensively assess the performance of the model.

By calculating these evaluation metrics, the predictive ability of the models can be quantitatively analysed and corresponding optimisation measures can be taken to address the limitations of different models. For example, for models with low accuracy, one can try to adjust feature selection or optimise model parameters; for models with low recall, one can try to increase training data or improve pruning strategies, etc.

5. Discussion and Outlook

This thesis explores the application of machine learning technology in NBA2K game and its limitations, and proposes corresponding optimisation methods and evaluation indexes. However, there are still some limitations in the current research, such as the difficulties of data collection and the selection of algorithmic models. Future research can further deepen the application of machine learning in game intelligence and develop more complex and efficient algorithmic models to enhance game experience and player satisfaction. At the same time, it is also necessary to pay attention to issues such as data privacy and security to ensure that the application of machine learning technology in games meets the requirements of relevant laws and regulations.

Summary:

This thesis proposes the method and idea of applying machine learning technology to NBA2K game, which can significantly improve the intelligence level of NBA2K game and make the game more challenging and interesting. The combination of machine learning and virtual reality games is an important direction for the future development of artificial intelligence games, which will bring players a more personalised and exciting game experience.

References:

[1]Zhang XG. On statistical learning theory and support vector machines[J]. Journal of Automation,2000(01):36-46.DOI:10.16383/j.aas.2000.01.005.

[2]Chen K,Zhu Y. A review of machine learning and its related algorithms[J]. Statistics and Information Forum,2007(05):105-112.

[3]Xuebing Yang,Jun Zhang. Decision tree algorithm and its core technology[J]. Computer Technology and Development,2007(01):43-45.

[4]Oduselu-Hassan Ph.D, Oladayo & Kenneth, Onyenike. (2024). Synergies Between Machine Learning, Artificial Intelligence, and Game Theory for Complex Decision-Making. Asian Research Journal of Mathematics. 20. 102-116. 10.9734/arjom/2024/v20i11863.

[5]Hu, Junjun & Li, Xiaoyan & Zhang, Yongkuo & Ai, Xiajie & Chen, Lei. (2024). The research on word game based on SIRS-ARIMA model and machine learning algorithm. 10.21203/rs.3.rs-4015563/v1.