Final Report

Dongjie Zhang,Noel Taterway

Abstract

NBA is one of the most popular sports organization and live sports game. It also has a large market value. Every year, NBA would have a NBA draft and it would give the teams to pick the rookie in the beginning of this year. This draft mechanism would give the teams with bad performance in last year the higher chance to get the better picks. Through researching this draft mechanism and doing data mining on the draftee's datasets, we could know how to improve the draft pick mechanism and help NBA develop better.

1. Introduction

It's known that NBA draft every year intends to give bad teams a chance to rebuild with draft picks, usually resulting in the last place team getting the first pick. However, that approach with a higher chance of getting the first pick does not make sure that the bad teams can select the best players always. for instance, Bill Russell with eleven NBA championships, kobe bryant winning five NBA championships, they two are not the first pick but both have a lot and are NBA great players.

The draft pick refers to the power of the NBA teams to pick rookies at the draft after the season ends. Each team automatically has a first-round pick and a second-round pick each year. First-round picks refer to the power of the NBA teams to pick rookies in the first round of the draft in the order they were scheduled. The most important first-round pick is the top pick, the second-place pick and the third pick. They are the first round of the first, second and third selection of rookie draft picks. The players who are picked by the champion are called "draft picks" and are ranked as "second place show" and "third talent show" by the players in the second place and third place. The winning of the top three draft picks ("First Prize", "Second Prize" and "Third Prize") must be decided by lot. The teams that qualify for the draw are the 14 teams that did not qualify for the playoffs that year. Usually the league will be in the playoffs when the fighting is in full swing the 14 teams pick the draw draw. After the champion pick out, with the same approach will also draw the second place to sign and the third sign. Only the teams that have been drafted in any of the top three picks must not participate in the next round of draws.

According to the rules of the league, the top three picks (ie "First Prize", "Second Prize" and "Third Prize") must be determined by lot. The teams that qualify for the draw are the 14 teams that did not qualify for the playoffs that year. Usually the league will be in the playoffs when the fighting is in full swing the 14 teams pick the draw draw. The host will draw 14 ping pong balls into a container when drawing lots. Each table tennis is marked with numbers 1-14. Draw four of these numbers when drawing lots. Excluding 11-12-13-14 these four combinations and the order, a total of 1,000 kinds of permutations and combinations. The league with the worst record has a combined 250 draw (or 25% chance) and the best lottery in the lottery is only 5% likely, and the better the record, the better less money you have, the lower your chance of winning a champion pick. Draw lots if you draw the same four balls as your own. If the league's worst record is not drawn to the lottery, then it is the second-lowest pick for the draw, and if not, then the third-worst team in the league will continue

pumping ... and so on until Draw the champion so far.

After the champion pick out, with the same approach will also draw the second place to sign and the third sign. Only the teams that have been drafted in any of the top three picks must not participate in the next round of draws. For example, if the team that picks the lottery is the penultimate team in the league and the team with the second best record wins the second worst record in the league, Four rotten team, then the three teams do not enjoy the right to draw in reverse order, but because of the remaining teams in accordance with record flashback distribution. The team with the worst record gets the No. 4 pick and the team with the fifth lowest score gets the No. 5 pick and the team with the bottom 6th record gets the No. 6 pick and so on ...

dBy researching the NBA draft pick mechanism, we could resolve many problems like:

1) Is having the highest possible pick always the best option?
2) How much did the rookie actually contribute to the team?
3) How much better did the team do after the draft pick?
4) Which picks result in the best contributions and players?

We could find the factors that influence the teams and players' performance in this research. After that, these findings could help us to improve the draft pick mechanism so that it can ensure that the weakest team in last season could get a relative better performance in the next season. Moreover, the first several picks do not actually mean that they are suitable for the last several teams. So by resolving the above problems, the draft pick mechanism could be improved and it can promote the performance of players and teams if every team could get the best option.

2. Related work

In 1988, only one of the three bottom teams got the chance to choose the top three. The Clippers with the lowest winning percentage finally got the draft of the year when he was forward Danny Manning from Kansas. 1989 to the present: Only two rounds of draft picks in the early draft, the team will have been selected until no available. The 1960 draft totaled 21 rounds. By 1974, the draft was stable at 10 rounds. By 1985, it was reduced to seven. In 1989, the National Basketball Association limited the draft to only two rounds to give unselected players a chance to follow the team's trial. Eight rookies from all walks of life have participated in the All-Star Game since the start of the two-round draft pick: Jazz guard Jeff Horona (1992), Pistons forward Dannis Rodman, Lakers guard Sadelli Kellogg-Sebastian (1995), Raptors center Anthony Davis (2001), Bucks guard Michael Reid (2004), Wizards guard Gilbert Arenas (2005), Spurs Defender Manukino Billy (2005) and supersonic forward Rashard Lewis (2005). In 2003 Detroit Pistons center Ben Wallace and Indiana Pacers center Brad Miller were the only players to be unselected but eventually became All-Star. From 1990 to the present: a favorable drawing system In October 1989, the coalition's management further improved the drawing lottery system. The 1990 draft, due to the development of the league, a total of 11 teams participated in the draw. The new rule states that the teams with the worst record in the regular season have 11 (11/66) chances to win the draft pick, and the team with the second worst finish record 10 (10/66) chances to win the draft pick. The teams that participated in the draw, who failed to play in the playoffs but the best record, had a 1 (1/66) chance of getting a draft pick. Orlando Magic hit Grand Canal, won the No. 1 pick for two consecutive years. In 1992, the Orlando Magic finished second in the

league with 21 wins and 61 losses. They selected the year's draft pick to come to lsU center Shaq-O'Neill. O'Neal helped the Orlando Magic get more than 20 wins, but still missed the 1993 playoffs. In the second year of the draft, Magic still selected Chris Weber, a power forward from the University of Michigan, despite having only a 1 in 66 pick. Magic immediately traded Weber to the Golden State Warriors for three draft picks and the then third annual pick from Memphis guard Penn Finney Hardaway. Union management revised the draft rules again in 1994, increasing the chances of the weakest team being drawn to the top three rookies and reducing the odds of a better-than-expected record. This change, will be the worst record drawn team draft picks from 16.7% chance of the original to 25%, while drawing the team's best team in the draft pick the odds of dropping from the original 1.5% To 0.5%. Under this provision, 14 table tennis balls marked with numbers (1 to 14) are placed in a bucket. Draw the ball from the 14 ball out of four. Regardless of the order of arrangement, the 4 numbers are combined in total by 1001. Prior to the draw, organizers will be 1000 combinations, according to the regular season record, points to the ballot team. After that, four ping pong balls will be pulled out to form a combination. The team that owns this combination has the No. 1 pick. After that, the ball is returned to the bucket and the entire process is repeated to determine the No. 2 pick and the No. 3 pick. 1996-2003: 13 teams in the draw In 1995, the NBA added two new teams, the Toronto Raptors and the Vancouver Grizzlies (now the Memphis Grizzlies), and the number of teams participating in the draw increased from 11 To the present 13. From 1996 onwards, the worst-of-the-record team is still 25% (250 combos) with the number one pick; the bottom second (20%, 200 combos) to the bottom sixth (6.4% 64 combos) (4.4%, 44 kinds of combinations) get the top pick the overall probability and the previous ratio unchanged; the last eight (2.9%, 29 kinds of combinations) to the countdown The probability of the twelfth (0.6%, six kinds of combinations) increased a little bit. Reciprocal thirteenth (0.5%, 5 combinations) the same chance of getting top rookie. The teams with the same record share the same percentage of chance, and if the sum of odds is the number of points, the two teams draw lots to decide who gets an extra percentage. 2004 to the present: 14 teams to participate in the lottery Because of the Charlotte Bobcats, the 2004 draft draw a total of 14 teams to participate. However, the Bobcats draft pick in 2004 was fixed for the fourth place, so in the lottery and no chance to get other pick rookie.

3. Dataset

   The Dataset we use in this task is the NBA draftees' statistic dataset from 1976 to 2015 and the corresponding team performance dataset. We put some statistic graph in the following pictures (Figure 1).

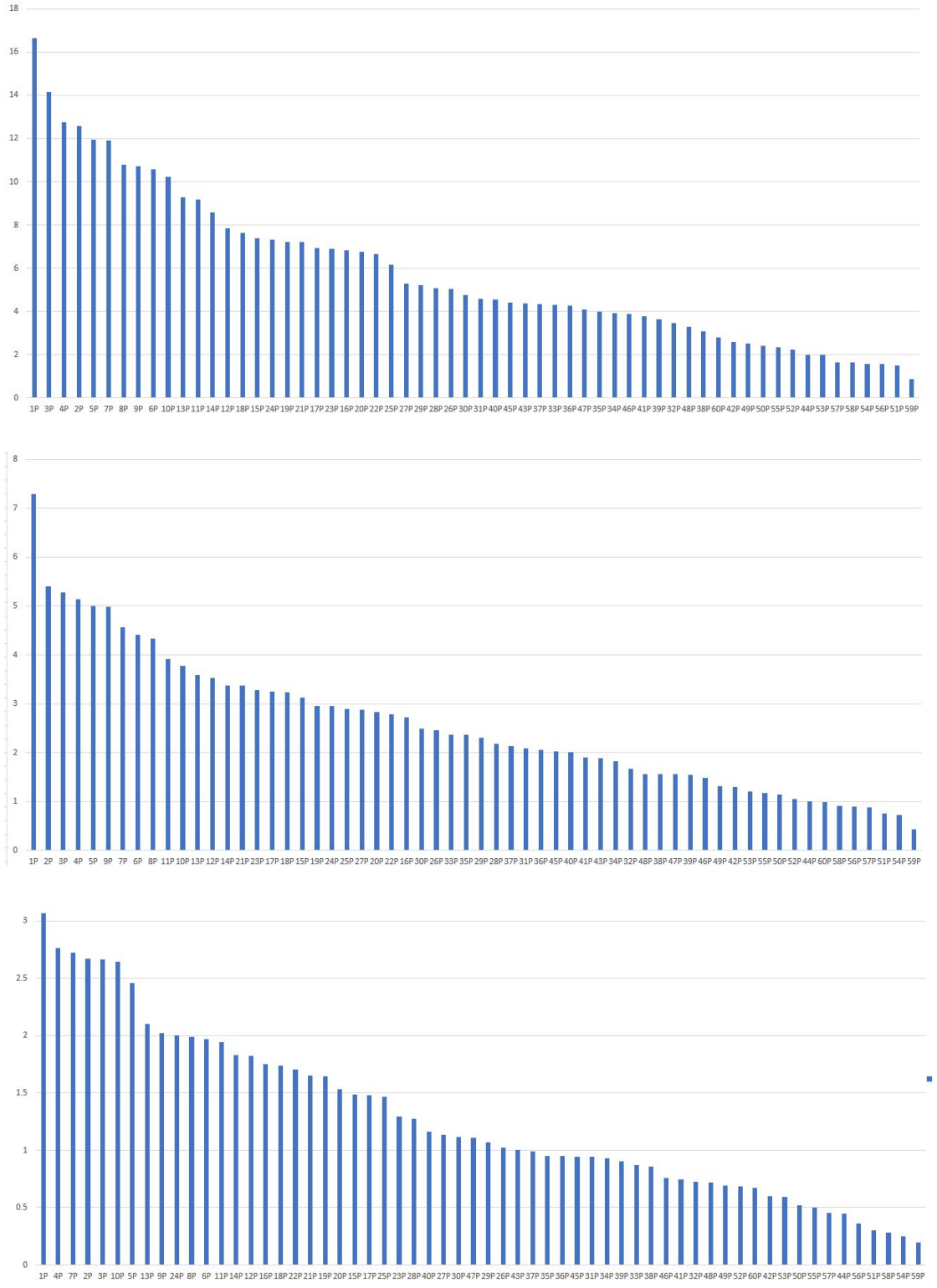


Figure 1 Some statistic data of the draftees

The first graph is the average points per game for the draftees, the second one is the average rebounds per game, and the last one is the average assists per game.

4. Main technique applied

4.1 Research Method

4.1.1 Logistic Regression

Logistic regression is such a process: in the face of a regression or classification problem, establish a cost function, then iteratively solve the optimal model parameters through the optimization method, and then test to verify the quality of our solved model.

Although Logistic Regression has a "regression" in its name, it is actually a classification method and it is mainly used for two classification problems (that is, there are only two kinds of output, representing two categories respectively).

In the regression model, y is a qualitative variable, such as y=0 or 1. Logistic method is mainly used to study the probability of certain events.

The function we used in the Logistic Regression is usually the Sigmoid function, it is a step function and it is shown in Figure 2.

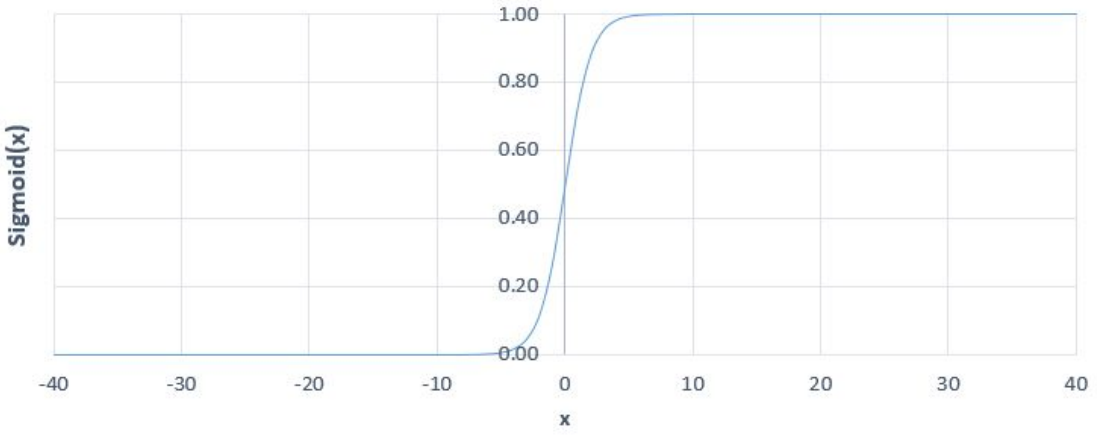


Figure 2 sigmoid function

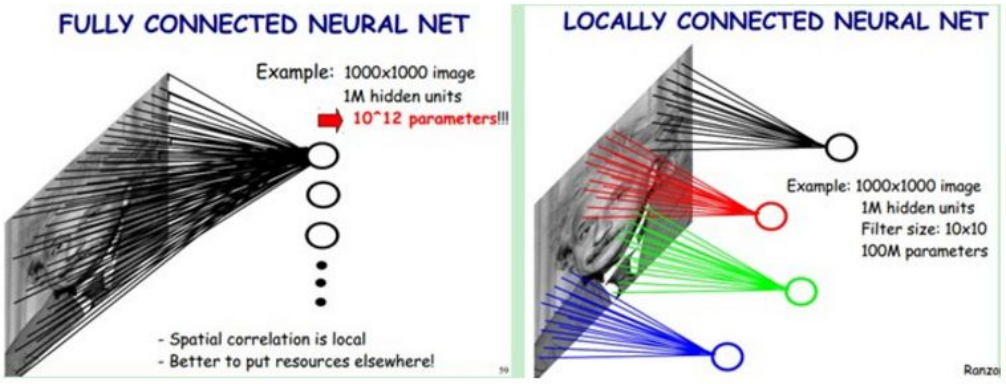
4.1.2 Convolution neural network

Convolution neural network (CNN) was proposed by Yann LeCun of New York University in 1998. CNN is essentially a multilayer perceptron, the key reason that it became a successful model is its local connection and weight sharing structures. These two designs of CNN reduce the number of the weights of the network and make the optimization of network easier. On the other hand, they can help to reduce the risk of overfitting. Its weight sharing network structure makes it more similar to the biological neural network, reduces the complexity of the network model. The advantages are more obvious when the input of the network is multi-dimensional image, because the image can be directly used as the input of the network, thus avoiding the complex feature extraction and data reconstruction process in the traditional recognition algorithm. There are many advantages in two-dimensional image processing using CNN, such as the network can automatically extract image features including color, texture, shape, topology, etc. CNN's structure also has good robustness in image processing, because this network structure is highly invariant to translation, scaling, tilting, or other forms of deformation.

The main three features of CNN are:

(1) Local (sparse) connection

In the traditional deep neural network, one layer and the next neighbouring layer are fully connected between the neurons. But in the convolutional neural network, neuron node between the layers is not fully connected. According to the local spatial correlation, neurons of each layer is only connected with the neuron nodes which are subparts of the upper adjacent layer, this is called the local connection. In this way, it greatly reduces the parameter size of the neural network architecture. In Figure , the left image is the fully connection, right image is the locally connection.



Local connection structure

(2) Weight sharing

In the convolutional neural network, each convolution layer has several convolution kernel, and each convolution kernel would filter the input data repeatedly. After convolution of the input data, the output of convolution layer would formed into several feature map, and these feature maps can be seen as the result of extracting the local features of the image. Each convolution kernel shares the same parameters, including the same weight matrix and bias terms. The advantage of shared weight is that the feature of the image is extracted without considering the location of the local feature. Moreover, weight sharing provides an efficient way to reduce the number of parameters of the convolutional neural network model greatly as well.

(3) Pooling (Sub-sampling).

After obtaining the image feature maps by convolution, these features should be used for classification. The classifier can be trained with all the extracted feature data, but this usually results in a huge amount of computation. Therefore, after obtaining the convolution feature of the data, the dimension of convolution feature would be reduced by the maximum pooling method (sometimes it could be average pooling or other pooling method). The convolution feature is divided into several disjoint regions (n*n), and the maximum (or average) feature of these regions is used to represent the convolution feature after dimensionality reduction. These dimensionality reduction features are easier to process and classify.

These two operations (convolution and pooling) are performed on two different layers in CNN, called the convolution layer and the pooling layer, respectively. In general, the convolution layer is the feature extraction layer, each neuron is connected with a subpart of input data feature, and the output of each

convolution layer is a feature mapping (local feature extraction); pooling layer is a layer of general sampling (i.e. pool), and it can be seen as another kind of feature extraction. Each layer usually has a number of feature maps, the weight parameters in one feature map are shared. And then each layer is connected with the latter layer. The convolution layer and the pooling layer are connected continuously, and the features of different levels are extracted.

4.1.3 LSTM + Attention mechanism

Worst teams sometimes do not get a good lottery result, which is why setting up a data set is especially important if you get better results through scientific analysis and statistics. Here, I suggest collecting all the draft data, refined to each team's draft results every time.

For the rookie, you need to establish a formula to rank rookies to ensure that each player is selected the same chance. Here to exclude the impact of the year and other objective factors. Second, the collected data should have the appropriate procedures to verify the reliability of the data. Then the collected data analysis, re-determine the order of arrangement. Systematicize vintage and team selection of players and determine the best scenario to determine the choice that will lead to the best player. Reordering the picks based on the data (2, 5, 1, 10, etc.) rather than using a fixed pick order, instead using the existing data and rankings to determine the draw order. You can also find other modes, change the order of generation and so on.

More specifically, we need to collect the rookies' performance data in each game of their first season in the NBA. We also need to collect the whole team's performance data in each game of the corresponding season. Moreover, the rookies' personal features and each team's features are also needed. The rookies' personal features include their height, weight, position, shooting-average, etc. As for the whole team's features, it could be implicated by the combination of the whole players' features in this team. In total, the data we used in our research could be divided into two sub parts: 1) Performance data; 2) Feature data. In addition, because our proposed model is a supervised model, we also need the ground truth (labels). The labels here could be a 2-classes label, 0 and 1 respectively. In which 1 denotes the team has a better performance than last season, 0 denotes that the team has worse performance than last season. The criterion of performance is determined by the rank of each team in NBA. For instance, if Magic team gets 10^{th} rank in this season and it got 9^{th} rank in last season, we consider Magic has a worse performance this season, so its label is 0. These two kinds of data and corresponding labels should be used in our proposed model to determine which player (rookie) is best suitable for each team.

Because the rookie performance and team performance data is a time-sequence relevant data, which means that it has temporal dependency. We could discover the history information inside the whole performance dataset to help us to make a better model. LSTM could utilize the time-sequence dataset and analyze the historical information underlying it. In this way, LSTM could be helpful in solving the rookie matching task.

Team feature data has its spatial distribution, in other words, only some special features of players could influence the team's performance, it could be the players' height or shooting average or any other factors. It would depend on the team's game style and players' constitution. So different team feature should

have different weights in causing the performance better or worse. In order to analyze the spatial distribution information, we would use the stack attention mechanism to find corresponding distribution inside the team feature.

In this project, we proposed to use an Attention-based Multi-Layer Perceptron (MLP) model to resolve the rookie matching problem. This new model contains CNN, LSTM, stack attention mechanism and MLP. The overview structure of my proposed model can be seen in Figure 3. First, team performance and rookie performance data would pass through a LSTM network to generate the final performance feature. For team feature, a pretrained CNN: resnet-152 would be used to get the its high level abstract representation. After we get the feature of performance and team, An stack attention mechanism would be applied on them to extract the distribution information. Finally, all these features would be sent into a MLP to get the output.

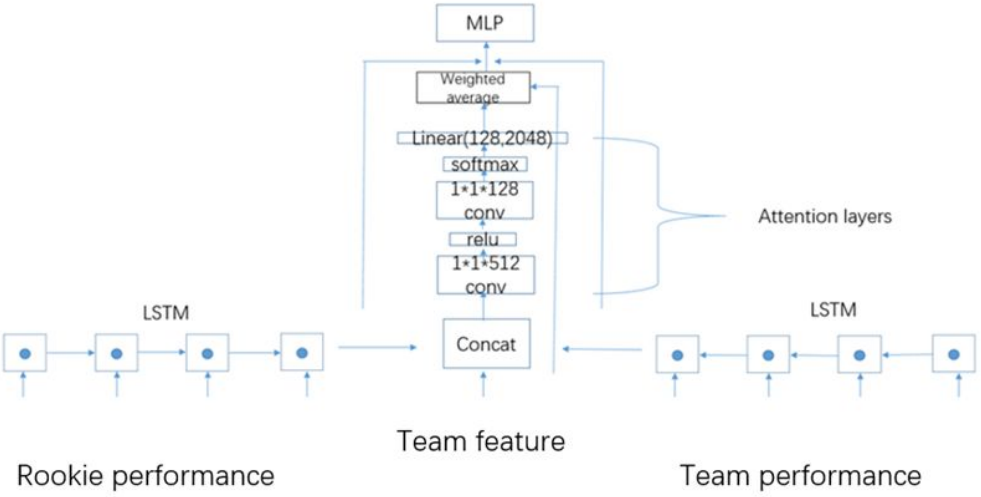


Figure 3 The overview structure of the model proposed in my project

The Attention-based Model is actually a measure of similarity. The more possilble the current input is relevant to the target state, the greater the weight of the current input would be, in other words, it indicates that the current output depends more on the current input. Strictly speaking, attention is not a new model, but merely adding the idea of attention to the previous model, so we always called it Attention-based Model or the Attention Mechanism. Attention mechanism is usually realized by a multi-layer neural network, and it would compute a attention distribution vector.

LSTM is a special case of Recurrent neural network (RNN) which internal state can show the dynamic timing behavior. Different from the feedforward neural network, RNN can handle time sequence data using its internal memory. It contains a cycle of neural network, and it has the explicit modeling ability of time sequence data by adding the connection of hidden layer across time steps in neural network. And in this way, it allows the long-term persistence of information.

After the model is designed completed as above, we could use the extracted

features to train this model. The team features would be processed in the attention-based model and generated an output vector contained the spatial information of static analyses, and the performance data would be transferred into another vector contained the temporal information of dynamic analyses. Finally, these two vectors would be concatenated and pass through a classifier to predict the rookie is suitable for the team or not.

4.2 Tools

Python

Python, a widely used high-level programming language, is a general-purpose programming language created by Guido van Rothsum. The first edition was published in 1991. Can be seen as an improvement (adding the advantages of some other programming language, such as object-oriented LISP). As an interpreted language, Python's design philosophy emphasizes code readability and concise syntax (especially using space indentation to divide blocks of code rather than braces or keywords). Compared to C++ or Java, Python allows developers to express ideas in less code. Whether it is a small or large program, the language tries to make the structure of the program clear.

Pandas

The Python Data Analysis Library or pandas is a tool based on NumPy that was created to solve data analysis tasks. Pandas incorporates a large number of libraries and some standard data models, providing the tools needed to efficiently operate large data sets. Pandas provides a large number of functions and methods that allow us to process data quickly and easily.

Numpy

The NumPy is an open source numerical computing extension of Python. This tool can be used to store and process large matrices much more efficiently than Python's own nested list structure (this structure can also be used to represent matrices)

Jupyter Notebook

The Jupyter notebook (also known as the IPython notebook) is an interactive notebook that supports running more than 40 programming languages.

Pytorch

PyTorch is a deep learning tensor library optimized using GPU and CPU.

4.3 Evaluation

There are many different parameters inside this model, like the number of factors in the rookie feature and team feature, the time-steps of LSTM, etc. We need to test different configuration in the future experiments.

5 Key result

1	13	17	45	41	57
3	11	20	31	34	50

5	14	19	36	39	55
2	18	22	35	32	53
4	12	29	37	48	44
9	24	25	43	38	58
7	15	26	33	42	56
8	21	27	40	49	54
10	23	28	47	60	51
6	16	30	46	52	59

The models used in this task can help us to predict the performance of the draftees, including their several statistic results. And the attention based model is constructed using pytorch and it could be trained using the GPU. It is a relevant complex model and we predict that it could get good performance and results in team –rookie matching.

Using this data, you can found out which type of player you can draft at your pick at a better price. For example, the 24th pick on average can give you the same amount of assist per game as the 7th pick. The 45th pick can give you the same number of rebounds as 28th pick. The 48th pick can give you the same number of minutes as the 32nd. These stats when you look at in context have to be the same player. If you pick a point guard at 24 you can possibly get a passer as good as a 7th pick, but that won't work if you draft a center whose man job is to get rebounds.

6. Application

What this truly tells us is the true probability of the pick being a bang or a bust. Our data will not be consistent for every draft, but it is just showing the probability of how good this pick can truly be and what they can strive at. Of course, there has been number #1 picks who have not been good, but majority of them are good players. Pick 3 & 5 are better than pick 2 historically because more 3 & 5 picks have played well in their careers than pick 2. We are not saying that if you have the 31st pick you should trade it for the 45th pick because of the true order. Being the 31st pick you should have more confidence, due to the higher pick by probability and chance on paper. If you are the 45th pick though, you should have extreme confidence because even though you are a later pick, the data tells your pick will pan out.

The knowledge we gained from this task can help the NBA improve their draft pick mechanism, and it also can help team to find what players they really need. Moreover, draftees also benefit from it because they could have better performance in their games.