Name: Chun-Chi Huang
Course: Data Modeling
Professor: Michael McKee
Date: 03/19/2018

Data Science Applies to Basketball Strategies

When I was a junior high student, I started playing basketball. Over the years, I have spent a lot of time improving my skills in basketball. Even though I am no longer playing competitively, I still desire to sharpen not only my offensive capabilities but also my defensive abilities. I joined both the senior high school team and university team where I learned plenty of strategies. I have also been watching NBA games since 1992.

One game that I participated at senior high school has interested me for several years. I remembered that my team lost by thirty points to a top team. After that, we only had one week to prepare for another game against the same team, and it was not enough. Our average height was ten centimeters shorter than the rival team. In addition, they also had many star basketball athletes, while our team members only came from normal schools that just focused on entering academic universities. In every perspective, my team hardly had a chance to win the game. My coach also understood this situation, so he changed his training strategy for our team. He required us to watch the game that we lost the previous week, and he tried to find a way to beat our opponent. Finally, we changed our defensive tactics from a 2-1-2 to a 1-3-1 strategy and practiced this tactic for one week. Surprisingly, we ended up winning this critical game by thirty points. That was incredible, and even now it is still a mystery to me.

For these reasons, I am very curious about the strategies in basketball games. Different strategies can affect the games tremendously with the same players, like the game mentioned above. *Moneyball* also taught me that data science has impacted sports like baseball significantly. In this paper will discuss some research introducing data science into basketball games. The first research paper is about the influence caused by different offensive strategies, and the second research paper uses link prediction to analyze the offensive tactics of basketball games. In addition, the third research paper discusses spatio-temporal learning in basketball offensive strategies.

The first research paper is “The Benefits of Forcing Offensive Basketball Players to their Weak Side” by Bartholomew and Collier (2012). There is a motto in basketball that “Defense wins championships”. This motto demonstrates that defense is the most important element of winning games. It can also apply to competitive sports at any level. The research question of this paper is what benefits come from defensive players forcing offensive players to use their weak sides. Before investigating this topic, the goal of defense and how to measure the defensive efficiency in basketball must be understood. The goal is to fight with the offensive players and make them have lower total points or lower field goal percentage. Based on this concept, this paper aims to find the relationship among “Weak Side (WS)”, “Strong Side (SS)”, some defensive indicators, and some offensive indicators. In this paper, Bartholomew and Collier (2012) use digital video recording software, Gamebreaker, to collect the performance data by player and by team. The data comes from NCAA Division I and II college basketball games in which players typically depend on their strong sides, and their strong sides are only left or right hand, not both. This means that these players have higher performance with their strong sides than with their weak sides.

According to the analysis of correlations of basketball variables for one university team, there are some explainable results shown as follows:

1. The correlations between of the number of WS, total opponent points (TOP), field goal percentage (FG%), and 3-point field goal percentage (3FG%) are higher than the correlations between the number of SS, TOP, FG%, and 3FG%. This means that if offensive players go to their strong sides (SS), their performance indicators including TOP, FG%, and 3FG% increase more than when they go to weak sides (WS).
2. The correlations between defensive rebounds (DR), TOP, FG%, and 3FG% are negative. This indicates that when DR decreases, TOP, FG%, and 3FG% of offensive players will increase.
3. The correlations between the number of WS, DS, offensive assists (OA), and DR are lower than the correlations between the number of SS, DS, OA, and DR. This demonstrates that players perform DS, OA, and DR greater with their SS than with their WS.
4. The correlation of forced turnovers (FTO) is higher with WS than SS. This shows that when players go to their WS, their FTO will increase.
5. According to the following equation, DS plays the most important role in TOP, and DR is the second most valuable factor. In addition, forcing players to their SS has less worth than the previous two defensive indicators, but it still decreases TPO.
   $$\mathrm{TOP} = 41.1541 - 0.229118 * \mathrm{WS} + 0.468505 * \mathrm{SS} - 1.74644 * \mathrm{DR} - 2.03136 * \mathrm{DS}$$

The second research paper is "Analysis of Offense Tactics of Basketball Games Using Link Prediction" by Zhang, Hu, and Liao (2013). This paper focuses on offensive behaviors of the rival team and attempts to find out the cooperation between different players who collaborate to score. Accordingly, this paper uses graph-based link prediction method to analyze the game records for basketball teams' offensive tactics.

In this paper, it considers offensive tactics as relationship between rival players. The relationships of executing ball-passing, making the final action, and showing the result constitute an offensive chain. For example, in Fig. 1, the first offensive chain is that player 1 passes the ball to player 5, and then player 5 passes the ball to player 2. Later, player 2 makes the 3-points shoot, and the result is scored.

This approach includes two main steps. The first step constructs a Basketball Analysis Graph (BA graph) from the data of offensive chains collected in the basketball games. The second step introduces link prediction into the BA graph to determine the important links. In general, the particular offensive tactics analysis may use the rival team's previous three games, and the number of offensive chains can reach up to 300. Additionally, Fig. 2 shows an example of a BA graph based on Fig. 1. The numbers labeled on the lines are the times of ball-passes.

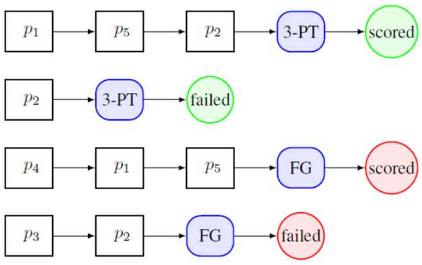
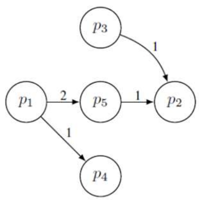


Fig. 1. Offensive chains



(a) BA graph

Node	Action	Result
p_2	3-PT	scored
	3-PT	failed
p_4	FG	scored
p_5	FG	failed

(b) Action/result information

Fig. 2. BA graph of the offensive chains

Next, Zhang et al. (2013) constructs the BA graph using the "BuildGraph Algorithm" and the "AddVertex Algorithm" shown in Fig. 3. In these two algorithms the variables are defined as follows:

1. "V" is a set of players and their scoring information.
2. "E" is a set of ball-passing times from one player to another.
3. "G" is a BA graph with a vertex set "V" and an edge set "E."
4. "C" is a set of offensive chains.

Algorithm 1: BuildGraph(C)

```
Input: C = {c_i} — set of offensive chains (Formula (1))
Output: G = ⟨V, E⟩ — weighted directed graph
begin
    G ← ⟨∅, ∅⟩
    foreach c_i ∈ C do
        x ← 1st player p_i1 ∈ c_i
        if x ∉ V then
            AddVertex(x, G)
        foreach next p_ij ∈ c_i do
            y ← p_ij
            if y ∉ V then
                AddVertex(y, G)
            (x, y, w) ← (x, y, w + 1)
            x ← y
        end
        for action and result a, r ∈ c_i do
            add (a, r) information to x.table
        end
    end
    return G
end
```

Algorithm 2: AddVertex(v, G)

```
Input: v — a vertex
Input: G = ⟨V, E⟩ — a directed graph
Output: Graph G updated with v, edges (v, u) and (u, v)
        added for all u ∈ V
begin
    foreach u ∈ V do
        E ← E ∪ {(v, u, 0)}
        E ← E ∪ {(u, v, 0)}
    end
    V ← V ∪ {v}
    return G
end
```

Fig. 3. BuildGraph and AddVertex Algorithm

Name: Chun-Chi Huang
Course: Data Modeling
Professor: Michael McKee
Date: 03/19/2018

Once the BA graph is established, this paper applies the Katz index to find the most cooperative pair of players. The Katz index is the score of a pair of two vertexes shown in the following equation. The higher the score of the Katz index, the more critical pairs of players there are. Consequently, the pairs of players with high Katz scores play critical roles in their offensive tactics. Defensive players should emphasize these pairs of offensive players.

$$score(x, y) = \sum_{l=1}^{\infty} \beta^{l} \cdot |paths_{x,y}^{(l)}|$$

Finally, this paper uses this approach on one NBA game, with data from the first quarter of a Lakers vs. Blazers game. The result proves that the pairs of players with high Katz scores have higher offensive performance than others.

The third research paper is “Spatio-Temporal Learning of Basketball Offensive Strategies” by Chen, Liu, Wang, Chu, Tang, and Liao (2015). Chen et al. (2015) generates a systematic approach to establish unsupervised modeling for offensive tactics. Since offensive behaviors are dynamic player trajectories, this research divides basketball training data (trajectory sets) into clusters and studies their respective spatio-temporal model. Finally, the result creates a distinguishable model as well as improves the learning of the models. Analyzing these offensive trajectory sets helps the defensive team understand offensive team’s strategies, and then the defensive team is able to determine the weaknesses of the offensive team.

Fig. 4 displays the flow of the method for learning the spatio-temporal models from given video clips. First, to accomplish the desired clustering from video clips, this research represents the trajectory sets by pre-processing which uses a tracking by detection technique. Once the trajectory sets are obtained, this paper uses dominant-set clustering to find the best matching among trajectory pairs by dynamic time warping. The relationships of different clusters are shown in Fig. 5. Consequently, this research performs temporal and multi-sequence alignment and Gaussian mixture regression to finish the model learning. By using this model, defensive teams can predict the type of offensive strategies without depending on manual recognition.

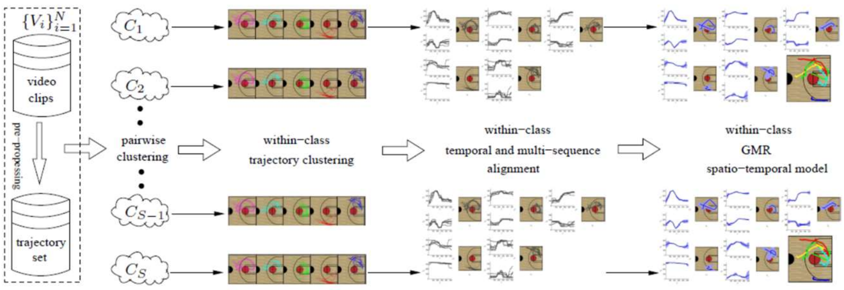


Fig. 4. Flowchart of learning the spatio-temporal models from given video clips

Name: Chun-Chi Huang
Course: Data Modeling
Professor: Michael McKee
Date: 03/19/2018

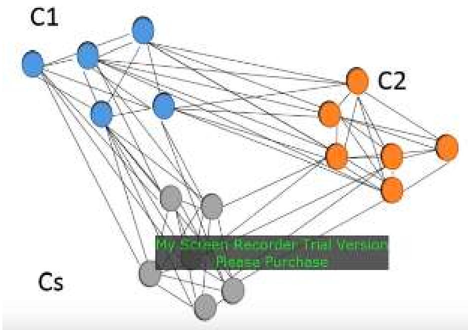


Fig. 5 The relationships among different clusters

In order to demonstrate the efficiency of this model, this paper performs the experiments on the NBA games. The result shows that this model reaches an accuracy of 89.1%. This excellent performance indicates that this temporal-temporal learning model can confidently distinguish and classify complex group behaviors, such as an NBA game.

The research discussed in the paper demonstrates that data science can greatly influence many aspects of basketball games. Correlation analysis can reflect the performance of basketball players on their weak sides and strong sides, and forcing offensive basketball players to their weak sides can lower their performance. Link prediction can prove that this method is able to find critical pairs of players in basketball games. In addition, Spatio-temporal learning mode can precisely identify rival teams' offensive tactics with 89% accuracy. Thus, professional basketball teams should invest more resources into data science analysis as it can help them achieve better performance in basketball games.

References

1. Bartholomew, J. T., & Collier D. A. (2012). The benefits of forcing offensive basketball Players to their weak side. *Journal of Multidisciplinary Research, Vol. 4, No. 2, Summer*, pp. 19-27.
2. Chen, C. H., Liu, T. L., Wang, Y. S, Chu, H. K., Tang, N. C., & Liao, H. Y. M. (2015). Spatio-temporal learning of basketball offensive strategies. *MM '15 Proceedings of the 23rd ACM international conference on Multimedia*, pp. 1123-1126.
3. Chen, C. H. (2015). Spatio-Temporal Learning of Basketball Offensive Strategies. https://www.youtube.com/watch?v=DGrzEleeL-c.
4. Zhang, T., Hu, G., & Liao, Q. (2013). Analysis of offensive tactics of basketball games using link prediction. *IEEE/ACIS 12th International Conference on Computer and Information Science (ICIS)*, pp. 207-212.