**Summary**

**Have a wonderful soccer game**

本文利用图论，概率论和微积分的方法，利用数据分析和建立机器学习模型，为足球教练的阵容安排和球员训练提供策略。

This paper proposes a method, with graph theory, probability theory and calculus, to build machine learning models based on data analysis, which aims at providing strategies for soccer coach's lineup arrangement and players' training.

In this paper, graph theory, probability theory and calculus are used to analyze data and build machine learning models, so as to provide strategies for football coach's lineup arrangement and players' training.

Firstly，根据图论，在球员之间建立传球网络，并建立单次传球的价值评价模型，用于评价两两球员间传球的配合程度，即传球网络的边权。建立在一定时间范围内所有参与比赛的个球员的邻接矩阵，通过以个点的子完全图边权之和为排序关键字找出若干组优秀的元组合。同时建立基于时间尺度的价值模型，用于评价时间对传球效率的影响。

Firstly, the Pass Network Model can be established according to the graph theory, whose edge-weights are evaluation of coordination degree of each dyadic configurations. Pass Evaluate Index is designed for evaluate a single pass, and the summation of each pass can be defined as the edge-weights of PNM. For analysis, the adjacency matrix of N participating players within a period. Several outstanding M configurations can be found by the sort of M-element combination with the key of the sum of the sub-complete graph edge weights. What’s more, investigation of the influence of time on Pass Diversity depends on the constructed and approximate function of time and pass.

Firstly, based on graph theory, a passing network is established between players, and a single pass value evaluation model is established to evaluate the the cooperation degree of passing between two players, that is, the edge weight of the passing network. Establish the adjacency matrix of all N players participating in the game within a certain time range, and use the sum of the sub-complete graph edge weights of M points as the sort key to find several sets of outstanding M-element combinations. At the same time, a value model based on time dimension is established to evaluate the impact of time on passing efficiency.

Secondly，我们将反映成功团队合作的绩效指标划分为静态指标和动态指标。静态指标包括球员位置安排和球队阵型（line-up），我们建立球员赛季热点模型和球员分布模型。动态指标包括opponents，side，coach，passes，defence，attack and fail等。对经过数据清洗动态指标之间通过可视化进行相关性分析后，以净胜球分类作为比赛样本标签，以随机森林分类器作为机器学习的模型，用网格搜索调优参数，建立动态指标评价模型，进行交叉验证，达到了80%的准确率。

Secondly, performance indicators that reflect successful teamwork can be divided into dynamic indicators and static indicators. Static indicators include player position arrangement and line-up with which player season heatmap models and player position models can be established while the dynamic indicators include opponents’ strength, side, coach, passes, defense, attack and fail .etc. After visualized analysis of the correlation between the dynamic indicators extracted after data cleaning, and with the setting label by the goal difference, the random forest classifier is used as the machine learning model as the evaluation model of dynamic indicators. After the Grid Search used for tuning parameters, and cross-validation, the accuracy of the model achieving 80% approximately.

Secondly, we divide performance indicators that reflect successful teamwork into dynamic indicators and static indicators. Static indicators include player position arrangement and line-up. We establish player season hotspot models and player distribution models. The dynamic indicators include opponents, side, coach, passes, defense, attack and fail. etc. After visually analyzing the correlation between the dynamic indicators after data cleaning, the goal difference classification is used as the game sample label, and the random forest classifier is used as the machine learning model. We use the grid to search for tuning parameters to establish a dynamic index evaluation model, and performed cross-validation, achieving accuracy of 80%.

Thirdly，通过上述中建立的模型进行观察分析，我们着重研究静态指标对球队的胜利起到的关键作用，综合考虑球员位置和技术数据评价模型，建立不同球员在不同位置价值评价模型。通过模拟退火算法，优化11人排列组合的考虑，在局部最优解的父级搜索树进行搜索全局最优解，最终给出价值最优的首发阵容阵型图。此外我们还考虑以下三个次要影响因素：球员间默契度，主客场影响和教练安排。给教练提出的综合建议。

Thirdly, the study focuses on the role of static indicators in the performance of the team and establishes different players' value evaluation models in different positions which comprehensively consider the player’s position and technical statistical data evaluation. To optimize the value of 11-person permutation, we choose simulated annealing(SA) algorithm which searches the global optimal solution in the cousin points in the same minimized search tree as the local optimal solution has attained. The model finally gave the best starting lineup formation. In addition, we also consider the following three secondary factors: tacit understanding between players, home and away influence, and coaching arrangements. All analysis above can be concluded as comprehensive suggestion to the coach.

Thirdly, the team's structural strategies include static indicators and dynamic indicators. Through the observation and analysis of the model established, we focus on the role of static indicators in the victory of the team, comprehensively consider the player’s position and technical statistical data evaluation model, and establish different players' value evaluation models in different positions. Through the simulated annealing algorithm, we optimized the 11-person permutation and combination, and searched the global optimal solution in the parent search tree of the local optimal solution. The model finally gave the starting lineup formation. In addition, we also consider the following three secondary factors: tacit understanding between players, home and away influence, and coaching arrangements, and make comprehensive suggestion to the coach.

Finally，我们用哈士奇球队的案例来解释群体动力学。并用哈士奇球队建立模型得到的结论来说明如何设计更有效的团队，并对团队绩效指标进行补充。

Finally, we use the case of the Huskies to explain group dynamics. And use the conclusions obtained by the Huskies to build a model to explain how to design a more effective team and supplement the team performance indicators.