

DNN을 활용한 'KBO' 플레이오프 진출 팀 예측

박주혁[○], 이양재*, 한희창*, 전유림**, 문유진*

[○]한국외국어대학교 Global Business & Technology학부,

*한국외국어대학교 Global Business & Technology학부,

**한국외국어대학교 독일어통번역학과

e-mail: fqdsa123@hufs.ac.kr[○], didwo1256@gmail.com*, raphael102324@gmail.com*,

wjsdbfla98@hufs.ac.kr**, yjmoon@hufs.ac.kr*

Prediction of KBO playoff Using the Deep Neural Network

Ju-Hyeok Park[○], Yang-Jae Lee*, Hee-Chang Han*, Yoo-Lim Jun**, Yoo-Jin Moon*

[○]Division of Global Business&Technology, Hankuk University of Foreign Studies,

*Division of Global Business&Technology, Hankuk University of Foreign Studies,

**Division of German Languages and Cultures, Hankuk University of Foreign Studies

● 요약 ●

본 논문에서는 딥러닝을 활용하여 KBO (Korea Baseball Organization)의 다음 시즌 플레이오프 진출 확률을 예측하는 Deep Neural Network (DNN) 시스템을 설계하고 구현하는 방법을 제안한다. 연구 방법으로 KBO 각 시즌별 데이터를 1999년도 데이터부터 수집하여 분석한 결과, 각 시즌 데이터 중 경기당 평균 득점, 타자 OPS, 투수 WHIP 등이 시즌 결과에 유의미한 영향을 끼치는 것을 확인하였다. 모델 설계는 linear, softmax 함수를 사용하는 것보다 relu, tanh, sigmoid 함수를 사용했을 때 더 높은 정확도를 얻을 수 있었다. 실제 2022 시즌 결과를 예측한 결과 88%의 정확도를 도출했다. 폭투의 수, 피홈런 등 가중치가 높은 변수의 값이 우수할 경우 시즌 결과가 좋게 나온다는 것이 증명되었다. 본 논문에서 설계한 이 시스템은 KBO 구단만이 아닌 모든 야구단에서 선수단을 구성하는데 활용 가능하다고 사료된다.

키워드: 신경망(Neural Network), 인공지능(Artificial Intelligence), KBO(Korea Baseball Organization)

I. Introduction

야구는 모든 스포츠를 통틀어서 많은 데이터를 제공하는 스포츠다. 투수의 투구, 타자의 타격, 주루 등의 모든 행동이 그대로 기록이 되고 데이터가 된다. 모든 행동이 데이터가 되는 스포츠인 만큼 데이터를 어떻게 활용하는지에 따라 승부가 갈리기도 한다. 많은 야구팬이 경기에서 나오는 데이터에 관심을 갖고 해당 데이터를 토대로 다양한 인터넷 커뮤니티에서 열을 올리기도 한다. 대중으로부터 많은 관심을 받는 스포츠이면서 데이터가 승부에 영향을 주기 때문에 경기에서 발생하는 데이터를 바탕으로 많은 연구가 진행되고 있다.

본 논문에서는 실제 정확한 플레이오프 진출 확률을 예측하기 위해 데이터와 인공지능 커뮤니티인 캐글의 데이터 [1, 2]와 야구데이터 기록사이트인 스탯티즈 [3], 각 구단 홈페이지에서 2021년부터 1999년까지의 시즌별 개인 데이터를 추출한다. 그리고 팀의 평균 타점, 타자의 기록, 투수의 기록을 각 정규화 과정을 거쳐 수치화하고 각 시즌의 순위를 입력하여 인공지능망을 학습시킨다. 다음 단계로, 최적의 성능을 찾기 위해 각각의 변수, 은닉층의 수와 노드의 수를

조정하여 프로젝트를 진행한다. 또한, 모델의 성능을 측정하기 위해 실제 시즌인 2022시즌 결과를 예측해본다.

II. Preliminaries

1. Related works

승패 예측에 대한 연구는 은 특정 스포츠에 국한되지 않고 다양한 분야에서 진행되고 있다. KBO에서도 상황별 승리 확률을 제공하고 있으며, 이닝별 승부 예측에 대한 논문도 많이 존재한다 [5]. 야구 뿐만 아니라 축구, League of Legend 게임에 대해서 15분 이후의 데이터를 기반으로 승부 예측하는 연구 역시 존재한다 [4]. 그러나 대다수의 경우 특정 시점을 중심으로 한 것이고, 시즌 최종 순위에 대한 연구는 비교적 부족하다. 야구에서 가장 널리 알려진 파타고리안 승률 같은 경우 단순히 많이 득점하고 적게 실점하면 이긴다는 발상에

서 시작하여 득점과 전체 득실점의 비율을 통해 승률을 계산한다. 그러나 이는 전체적인 팀 순위를 증명하기 위한 자료로서는 부적절하다. 그 밖에도 OPS, WAR을 활용하지만 이 역시도 불완전하다. 그렇기 때문에 팀 스탯들을 기반으로 하여 시즌 도중 플레이오프 진출팀을 예측하고 운이 많이 작용한 것인지, 반등의 여지가 존재하는지를 파악할 수 있도록 한다.

2. Data sources

데이터셋은 Kaggle에서 제공하는 KBO Batting Data와 KBO Pitching Data를 이용하였다[1,2,3] 이 중에서 1999년 이후의 데이터들만 활용하였다.

III. The Proposed Scheme

1. Variables

승부 예측의 정확도에 영향을 크게 미치는 변수를 판단하는 과정을 거쳤다. 먼저 전체 변수를 넣어 프로그램을 실행한 후, Test Accuracy를 확인(약 0.78), 전체 변수 중 내용이 겹치는 부분을 제거, 그리고 weight가 음수인 변수를 제거하여 실행하였을 때 Test Accuracy 약 0.88정도가 나온 것을 확인할 수 있었다.

2. Data normalization

야구 데이터의 변수가 자연수의 비중이 높고, DNN프로그램 작동 시 변수의 범위가 0~1사이인 것이 좋기 때문에, 모든 변수를 가장 큰 수를 1 가장 작은 수를 0으로 두어, 각각의 값을 편차에 맞게끔 설정하였으며(X값, Y값)로 해당 연도 야구팀의 순위를 두었는데 플레이오프 진출 여부를 판단하기 위해 1위에서 5위까지는 0을 나머지는 1의 값으로 설정하였다.

IV. Experiment

손실함수 설정을 위해 ReLU, Sigmoid, Tanh, 그리고 Softmax를 직접 실험하여 비교하였다. 실험 결과 ReLU, Tanh, Sigmoid를 모두 사용하였을 때의 정확도가 0.8833임을 확인하고, 세 가지 함수를 활용하여 구성하였다. 다층 퍼셉트론을 활용하여 모델을 구성하였으며, 각 팀의 시즌별 기록을 바탕으로 입력층에 타자변인인 ‘RBI’, ‘OPS’, ‘runs’, ‘triples’, ‘stolen_bases’, ‘caught_stealing’, ‘bases_on_balls’, ‘GDP’, ‘SLG’와 투수변인인 ‘ERA’, ‘WHIP’, ‘shutouts’, ‘saves’, ‘P_hits’, ‘home_runs’, ‘intentional_walks’, ‘wild_pitches’ 값을 주었다. 이에 은닉층의 ReLU와 Tanh을 통해 가중치를 수정해나가며, Sigmoid 함수를 통해 프로야구경기의 플레이오프 진출 가능 혹은 불가능의 값을 output으로 도출하게 된다. 이를 위해 손실 함수는 binary crossentropy를 사용하였으며, optimizer는 adam을 사용하였다. 본 연구에서 설정한 Architecture는 Fig. 1과 같다.

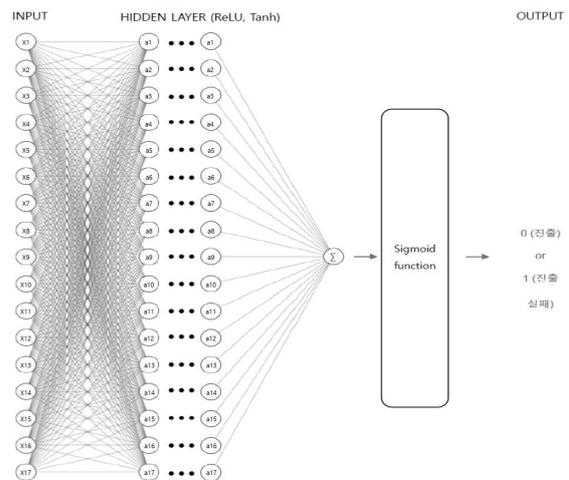


Fig. 1. Architecture of the KBO Playoff Prediction

실행 결과 테스트 정확도가 0.8833으로 대부분의 승부 예측에 성공했다. 이를 통해 작성한 프로그램이 프로야구경기의 플레이오프 진출 여부 예측에 유용하다는 결론을 도출할 수 있었다.

V. Conclusions

본 프로젝트에서는 시즌 데이터를 활용해 플레이오프 진출을 예측을 인공지능망을 이용하여 성공적으로 분석하였다. 또한, 가중치가 큰 항목을 파악해 선수보강시 구단이 집중해야하는 부분에 대한 정보를 제공할 수 있었다. 향후에는 팀 별 대진기록과 투수, 타자 개개인의 데이터를 활용해 경기의 승패를 정확하게 예측하는 연구가 가능할 것이다. 또한 보다 다양한 알고리즘을 결합하여 사용한다면 예측 정확도가 더 올라갈 것이다.,

REFERENCES

- [1] Kaggle, Batting Data, 2022. <https://www.kaggle.com/datasets/mattop/baseball-kbo-batting-data-1982-2021>
- [2] Kaggle, Pitching Data, 2022. <https://www.kaggle.com/datasets/mattop/korean-baseball-pitching-data-1982-2021>
- [3] Statiz, Team Data, 2022 <http://www.statiz.co.kr/league.php?opt=2022>
- [4] Si-Jae No, etc., “Prediction of League of Legends Using Deep Neural Networks,” Proceedings of KSCI Conference 2021, pp. 1-4, 2021.