

인공지능을 이용한 스포츠 상황 분석 서비스 : 축구의 기대 득점을 중심으로

김진섭^o, 김민준^o, 이관형^o, 윤용수^o, 문재현^{*}

^o홍익대학교 컴퓨터공학과

^{*}한국기술거래사회

jinsob@gmail.com, mjun@naver.com, rhks gud135@naver.com, fheodsla@naver.com, smjhoon@gmail.com

^{*}교신저자(Corresponding Author)

Sport Situational Analysis Using Artificial Intelligence : Focused on Football Expected Goal

Jin Sob Kim^o, Min Jun Kim^o, Kwanhyeong Lee^o, Yongsoo Yoon^o, Jaehyun Moon^{*}

^oComputer Science, Hongik University

^{*}Korea Technology Transfer Agents Association

요 약

스포츠팀 운영에 있어서 경기 중 상황에 대한 통계와 분석을 통해 좋은 성과를 내는 것은 스포츠 야구 종목의 Sabermetrics를 통해 이미 증명된 바가 있다. 한편, 축구에서는 최근 들어 선수의 역량을 평가하기 위하여 객관적인 시각에서 슈터(Shooter)에게 주어진 기회, 즉 슈팅 상황을 바라보는 기대 득점(Expected Goal; 이하 xG)이라는 지표가 등장하였으나, 객관성이라는 평가 의도와 다르게 경기 내 각각의 슈팅 상황을 정의하는 것에 있어 축구 분석관들의 주관성에 의존하는 한계성을 지녔다. 본 논문은 xG를 산출하는 방식에 있어서 기존의 주관성을 배제하고 인공지능을 통해 상황을 정의하여 객관적인 평가지표를 지향하며 유의미한 통계적 수치를 지닌 xG를 도출함으로써 결과 위주의 분석만이 존재하던 축구 종목에 있어서 경기 중 상황에 대한 객관적인 판단 및 정의에 대한 방향성을 제시한다. 또한, 본 논문에서의 인공지능은 국내 K리그 슈팅 데이터를 통해 학습되어 K리그 내 전략적인 상황들에 대한 특화된 xG를 도출하며, 이를 웹을 통해 K리그 내 선수 개개인에 대해서 시계열, 상대 팀, 슈팅 위치별 그래프로 시각화하여 제공하는 시스템을 구축함으로써 K리그를 기준으로 선수에 대한 평가 및 경기 운영에 기여할 수 있는 기대 득점 분석 서비스를 제공한다.

1. 서론

야구에서는 Sabermetrics를 이용한 경기 상황에 대한 분석이 성공적으로 운영되고 있다. 축구 종목에서도 역시 객관적인 시각에서 슈팅 기회를 고려하고 선수에게 주어진 상황, 즉 슈팅 기회가 득점으로 연결될 확률을 측정하는 기대 득점 [1]지표가 중요해지고 있으며, 최근 해외 리그에서는 이에 대한 활발한 분석 시도 및 연구가 진행되고 있다.

기대 득점(expected Goal; xG)이란 주어진 슈팅 상황에서 어시스트 유형, 골대와 공의 거리, 그리고 슈팅 각도 등 다양한 환경 요소들을 고려하여 주어진 기회가 골로 연결될 확률을 측정하는 지표로, 선수 개인의 기량에 의존하지 않으며, 슈팅을 시도한 경기 상황에 대해 과거의 동일, 혹은 유사한 슈팅 환경이라고 판단되는 슈팅의 결과들을 바탕으로 산출된다 [1].

2. 기존 서비스의 문제점

현재에 있어 경기 내 슈팅 상황에 대해 기대 득점을 산출

하는 방식은 객관적인 시각으로 선수에게 주어진 슈팅 기회를 평가하려는 의도와는 달리, 동일, 혹은 유사한 슈팅 환경을 정의하는 것에 있어 인적 자원, 즉 여러 축구 분석관들 간의 주관적인 판단 기준에 의존한다. 이는 슈팅 기회가 좋고 나쁨에 대해 책정되는 평가 수치가 필연적으로 객관적일 수 없음을 의미하고 기대 득점으로 선수 역량을 측정하는 것에 대한 불확실성을 야기한다.

이러한 문제점에 대해 본 논문에서는 기존 기대 득점 산출 방식이 가지는 주관성을 배제하고, 인공지능을 통해 슈팅 환경을 정의하도록 하여 유의미한 통계적 수치를 산출해내는 모델을 제안한다.

3. 연구 범위

[그림 1]에서와 같이 최근 해외 리그들에 대해서는 기대 득점을 팀, 선수별 평가지표로 사용하고, 구단 운영에 적용 및 공개하는 서비스들이 빠르게 증가하였다. 반면, 국내 K리그에 대한 기대 득점 분석 데이터는 현재까지 공개된 바 없

에 K리그를 중심으로 한 기대 득점 분석 시도의 필요성이 높다고 생각한다.



[그림 1 The English Premier League 2018/2019 시즌 구단별 기대 득점]

또한, 'K리그 데이터 포털'에서는 2015년도부터 현재까지의 K리그 경기들에서 발생한 슈팅 기록 정보를 공개하고 있으며, 본 시스템에서는 이를 크롤러를 통해 슈팅을 시도한 선수, 슈팅 결과에 대한 데이터와 매칭되는 경기 영상을 수집한 뒤 정제과정을 거쳐 K리그 내 슈팅 상황 기반으로 기대 득점을 산출하고 이에 대해 선수 개개인에 대해 다방면으로 시각화된 그래프를 사용자에게 제공한다.

4. 인공지능을 활용한 득점 예상 서비스

4.1 시스템 개요

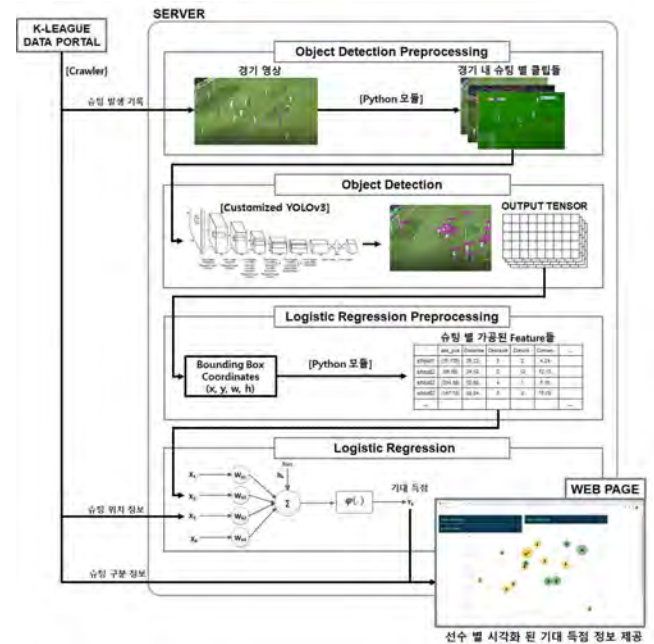
본 시스템에서 사용자는 웹을 통해 시스템에 접근할 수 있으며, 제공되는 페이지 내 UI 조작을 통해 선택된 선수와 팀에 대해 분석된 기대 득점 정보를 제공받을 수 있다. 제공되는 기대 득점 정보의 산출에 있어, 선수에게 주어진 슈팅 상황을 정의하는 과정에는 객체 추적 인공지능이 사용되었고, 슈팅으로부터 골로 연결될 확률 산출 과정에는 정의된 상황과 'K리그 데이터 포털'로부터의 슈팅 위치 데이터와 슈팅 결과로부터 학습된 로지스틱 회귀 모델 [3]을 활용했다.

4.2 시스템 구성도

본 시스템은 웹과 서버로 구분될 수 있으며, 서버는 K리그 경기 내 각각의 슈팅에 대한 기대 득점 분석 결과를 슈팅 구분 정보와 함께 DB에 저장하고 있으며, 웹의 선수 정보 요청에 응하여 프로필 정보 및 기대 득점 정보를 송신한다. 웹은 사용자의 조작에 따른 선수에 대한 정보를 서버에게 요청하고, 송신 받은 정보를 통해 동적으로 프로필 정보 및 기대 득점 그래프를 시각화하여 사용자에게 제공한다.

[그림 2]는 전체 시스템의 구성도이며 사용자가 제공받는 기대 득점 정보는 'K리그 데이터 포털'에서 크롤러를 통해 수집되는 슈팅 기록 정보에 기반한다. 수집된 슈팅 정보에는 경기 정보, 선수 정보, 슈팅 발생 시각, 슈팅 위치, 그리고 결과 정보가 포함되어 있으며, 서버에서는 이들을 통해 각각의 슈팅을 구분함과 동시에 기대 득점을 산출하기 위한 Ra

w Data로써 사용한다.



[그림 2 인공지능을 활용한 기대 득점 산출 로직 및 서비스 흐름도]

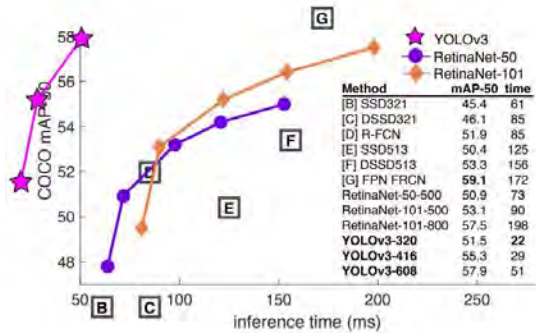
본 시스템의 기대 득점 산출 방식은 크게 세 과정으로 분리된다. 첫 번째는 슈팅이 발생한 경기 정보 및 슈팅이 경기 도중 발생한 시각 정보를 이용하여 경기 영상을 특정하고, 영상이 포함하고 있는 각각의 슈팅 클립 영상으로 분리한다. 두 번째는 각각의 슈팅 클립 영상으로부터 슈팅 상황을 정의할 수 있는 Feature들을 정제하는 과정으로, 'YOLO' 알고리즘 [2]을 이용한 영상 클립 내 객체 추적을 통해 슈팅 상황에 참여하는 객체들, 즉 선수들과 골대, 그리고 공의 상대적 위치 정보로부터 상황을 정의하는 정보를 추출한다. 마지막으로 세 번째는 객체 추적으로부터 추출된 상황에 대한 정의 정보와 슈팅 기록에 포함된 슈팅 절대 위치 정보를 Input Feature로, 슈팅 결과 정보를 Target Output으로 한 학습 데이터를 통해 학습시킨 로지스틱 회귀 모델이 예측하는 득점(1) 클래스에 속할 확률이 곧 기대 득점이 된다.

4.3 Object Detection 알고리즘 선정

슈팅 영상으로부터 상황에 참여하는 각각의 객체들의 상대 위치 정보를 파악하기 위한 Object Detection 알고리즘은 1-Stage Method, 2-Stage Method로 분류될 수 있다. 1-Stage Method 방식으로는 YOLO, SSD, RetinaNet 등이 있으며, 이미지 내 정해진 위치, 정해진 크기의 객체를 찾는 방식으로 이에 대한 전략적인 선택이 요구된다. 2-Stage Method 방식으로는 R-CNN, Faster R-CNN 등의 알고리즘들이 포함되며, 객체를 포함할 가능성이 높은 영역을 선택적으로 탐색하고 영역을 제안하는 Region Proposal Step과 후보군의 윈도우들을 취합 및 객체 탐지를 수행하여 객체가 존재할 확률이 가장 큰 윈도우를 반환하는 Fine-Tune Pro

posals & Final Prediction Step을 통해 객체 추적을 수행하는 방식을 일컫는다. 이 둘의 차이점으로는 1-Stage Method 방식이 2-Stage Method 방식보다 정확성은 떨어지지만, 실시간 탐지가 가능한 수준의 빠른 처리 속도를 보인다는 점이다.

본 시스템에서는 합리적인 시간 내에 과거의 K리그 슈팅 데이터를 처리하기 위해 1-Stage Method 방식의 'YOLO' 알고리즘을 채택하였으며, 그중에서도 기존 'YOLO' 알고리즘이 가지는 단점을 개선한 'YOLOv3' 모델 [2]을 사용하였다. [그림 3]은 Yolov3: An incremental improvement [2]에서 발췌한 대표적인 1-Stage Method 방식의 알고리즘들과 'YOLOv3' 알고리즘과의 처리 속도 및 예측 성능을 비교한 그래프이며 'YOLOv3'이 타 알고리즘 대비 처리 속도 면에서 압도하고 있음을 볼 수 있다.



[그림 3] 1-Stage Method 처리 속도 및 예측 성능 비교

4.4 데이터 수집 및 정제

본 시스템에서 기대 득점을 도출하는 과정에 필요한 데이터들은 과거의 K리그에서 발생한 경기 영상, 영상 내 슈팅 발생 시각, 슈팅 선수, 선수가 속한 팀과 상대 팀, 슈팅 위치, 그리고 슈팅 결과 정보이며, 이들은 경기 영상을 제외한 이들은 모두 'K리그 데이터 포털'로부터 크롤러를 통해 수집되며 경기 영상은 중계 영상 다시 보기 채널을 통해 수집되었다.

이후 Object Detection 전처리 모듈로 슈팅 발생 시각 정보를 이용하여 경기 영상들로부터 슈팅 상황이 발생한 순간의 짧은 영상 클립들을 추출하도록 함으로써 기대 득점을 환산하기 위한 서버 내 파이프라인의 인풋으로 활용한다.

4.5 슈팅 상황 정의

슈팅 상황을 정의하는 것에 있어 본 시스템에서는 'K리그 데이터 포털'로부터 크롤링한 슈팅 위치 정보와 Object Detection으로부터의 Output Tensor 내 감지된 객체들에 대한 Bounding Box Coordinates 중에서도 Bounding Box의 위치, 그리고 Class 정보를 활용한다. 크롤링한 슈팅 위치 정보는 슈팅이 발생한 절대적 위치를, 그리고 Bounding Box Coordinates들은 슈팅에 참여하는 객체 간 상대적 위

치를 표현하며, 해당 정보는 각 객체 간의 상대적 위치 벡터로써 표현 및 가공되어 최종적으로 절대 위치 정보와 함께 슈팅 상황을 정의하는 정형화된 Feature로 만들어진다.

상대적 위치 벡터를 통하여 객관화된 슈팅 상황을 정의하는 Feature를 생성한 방법으로는 객체 벡터 간의 외적 및 내적, 그리고 L2 Norm을 이용하여 슈팅 상황에 대해 참여하는 객체 속성들이 득점에 미치는 영향력을 분석하는 다섯 가지 척도들을 구현하였으며 이들은 다음과 같다.

*notations

Box center points : (x, y) , Box size : $(w, h) \in$ Bounding Box Coordinates

$(x_p^i, y_p^i, w_p^i, h_p^i)$: i -th Player's Box Coordinates

(x_b, y_b, w_b, h_b) : Ball Box Coordinates

$(x_{gp}, y_{gp}, w_{gp}, h_{gp})$: Goalpost Box Coordinates

수식 (1)은 공과 골대 객체 벡터 간의 L2 Norm, 즉 두 객체 간 유클리드 거리(Euclidean Distance)를 측정한다.

$$Distance(Ball, Goalpost) = \sqrt{(x_b - x_{gp})^2 + (y_b - y_{gp})^2} \quad (1)$$

(2)의 수식에서의 점 (A,B)와 점 (C,D)는 골대 Bounding Box의 각 꼭짓점 벡터들과 공 Bounding Box의 중심점이 만들어 내는 가장 큰 삼각형에서 공을 제외한 나머지 두 꼭짓점이며, 이에 대한 선수별 벡터의 외적 및 내적을 통해 해당 삼각형 내부에 위치하는 선수들을 추출한다. 즉, 공이 골문으로 향할 수 있는 모든 방향성 중 선수들의 위치로 인하여 가로막히는 경우(Obstacles)의 수를 계산한다.

$$Obstacles(Goalpost, Ball, Players) = \left| \left\{ (x_p^i, y_p^i) \mid \left((A_{gp} - x_b, B_{gp} - y_b) \times (x_b - x_p^i, y_b - y_p^i) \right) \cdot \left((C_{gp} - x_b, D_{gp} - y_b) \times (x_b - x_p^i, y_b - y_p^i) \right) < 0 \right\} \right| \quad (2)$$

수식 (3)은 공 벡터로부터 일정 Threshold보다 가까운 유클리드 거리를 가지는 선수 벡터들을 추출한다. 이들은 주어진 슈팅 순간에서의 공의 운동 방향성에 직접적으로 높은 영향력을 행사할 수 있는 벡터들, 즉 슈팅 상황에 있어 이 벡터들은 슈터 혹은 높은 영향력을 가진 방해자(Disturbance)임을 의미한다. Threshold 값은 로지스틱 회귀 모델의 평균 Loss에 대해 Fine-Tuning 과정을 수행함으로써 조정하였다.

$$Disturbance(Ball, Players) = \left| \left\{ (x_p^i, y_p^i) \mid \| (x_p^i - x_b, y_p^i - y_b) \| < threshold \right\} \right| \quad (3)$$

(4) 수식은 공 벡터에 대한 모든 선수 벡터들의 유클리드 거리(Euclidean Distance) 역수의 총합으로, 상황 내 각각의 선수들이 공에 대해 집중 및 마크할 수 있는 정도의 총합을 나타내며, 이를 통해 슈팅 상황 내 공에 대한 선수들의 밀집도(Concentration)를 반영하였다.

$$Concentration(Ball, Players) = \sum_{i=1}^N \left(\| (x_b - x_p^i, y_b - y_p^i) \| + \epsilon \right)^{-1} \quad (4)$$

마지막으로, 수식 (5)는 주어진 상황에서 공과 골대 사이 위치한 선수들이 공이 골문으로 향하려는 움직임에 대해 행할 수 있는 방해의 정도(Interference)를 측정한다. D 집합 내 원소들은 각각 골대 Bounding Box의 중심점, 그리고 각각의 꼭짓점 벡터들이며, Obstacle Candidates 집합은 공으로부터 각각의 선수들로 향하는 방향 벡터가 공으로부터 골대의 중심점으로 향하는 벡터와 직각 및 예각을 이루는 선수들에 대한 집합이다. θ (theta)는 공으로부터 각각 Obstacle Candidates에 대한 방향 벡터와 D 집합 내 각 원소로 향하는 벡터가 이루는 사이 각이다. 즉 Obstacle Candidates 집합에 포함된 임의의 선수가 가지는 득점 방해력(Interference)은 선수가 공이 골문으로 향하는 경로를 막아주기 위해 움직여야 하는 최단 거리에 반비례하므로 최단 거리에 대한 역수 형태를 취한다. 그리고 이를 Obstacle Candidates에 대해 공이 득점을 위해 목표로 하는 D 집합 내 도착 지점 원소들에 대해 각각 합산함으로써 주어진 상황이 가지는 득점 방해 요인을 측정하도록 하였다.

$$(\alpha, \beta) \in D = \left\{ \begin{array}{l} (x_{gp}, y_{gp}), \\ \left(x_{gp} + \frac{1}{2}(w_{gp} + h_{gp}), y_{gp} + \frac{1}{2}(w_{gp} + h_{gp}) \right), \\ \left(x_{gp} + \frac{1}{2}(w_{gp} - h_{gp}), y_{gp} + \frac{1}{2}(w_{gp} - h_{gp}) \right), \\ \left(x_{gp} - \frac{1}{2}(w_{gp} + h_{gp}), y_{gp} - \frac{1}{2}(w_{gp} + h_{gp}) \right), \\ \left(x_{gp} - \frac{1}{2}(w_{gp} - h_{gp}), y_{gp} - \frac{1}{2}(w_{gp} - h_{gp}) \right) \end{array} \right\} \quad (5)$$

$$\text{Obstacle Candidates}(\text{Goalpost}, \text{Ball}, \text{Players}) = \{(x_p^i, y_p^i) \mid (x_{gp} - x_b, y_{gp} - y_b) \cdot (x_p^i - x_b, y_p^i - y_b) \geq 0\}$$

$$\theta(\text{Ball}, D, \text{Obstacle Candidates}) = \arccos \left(\frac{(\alpha - x_b, \beta - y_b) \cdot (x_p^i - x_b, y_p^i - y_b)}{\|(\alpha - x_b, \beta - y_b)\| \| (x_p^i - x_b, y_p^i - y_b) \|} \right)$$

$$\text{Interference}(\text{Ball}, \text{Obstacle Candidates}, \theta) = \sum_{\text{Obstacle Candidates}} \frac{1}{\| (x_p^i - x_b, y_p^i - y_b) \| \sin(\theta) + \epsilon}$$

4.6 Logistic Regression 모델 학습

본 시스템에서는 주어진 슈팅 상황에 대해 해당 상황에서 골을 기대할 수 있는 정도를 예측하는 모델 파이프라인을 구축하고자 하였다. 따라서 상황을 정의하는 독립 변수들 간의 선형 결합을 통해 이에 대한 종속 변수, 즉 슈팅의 결과가 득점(1)인지 아닌지(0) 이항 분류 과정 중에 ‘득점(1)’으로 분류될 확률을 계산하는 이항 로지스틱 회귀 모델 기대 득점을 산출하는 것에 적합하다고 판단하였다 [3].

[표1]과 [표2]는 각각 로지스틱 회귀 모델의 기대 득점에 대한 학습 데이터 예시를 나타낸 표이다. [표1]은 Input Data X이고 각각의 행은 Instance로, 각각의 슈팅을 의미

하며, 열은 4.4와 4.6에서 가공하여 얻어진 슈팅에 대한 상황 정의 정보 값들을 나타낸다. [표2]는 4.4에서 얻은 슈팅 결과들로 [표1]에서의 각각의 Shooting Instance에 대한 결과 정답 값을 나타낸다.

	abs_pos	Distance	Obstacle	Disturb..	Concen..	...
shoot1	(35,125)	35.22..	5	2	4.24..	
shoot2	(86,56)	24.18..	2	12	32.13..	
shoot3	(204,58)	32.98..	4	1	5.35..	
shoot4	(167,75)	38.64..	5	4	15.78..	
...						...

[표 1 학습 데이터 Input X 예시]

	Y
shoot1	0
shoot2	1
shoot3	1
shoot4	0
...	...

[표 2 학습 데이터 Target Output Y 예시]

5. 결론

본 논문에서는 인공지능을 통해 슈팅 상황에 대해 정의하고 그에 대해 평가하는 로직을 구현하였으며, 이 로직을 통해 K리그 내 경기들을 학습하여 이를 기반으로 선수 개개인의 기대 득점을 분석하고, 이를 프로필과 함께 시각화하여 사용자에게 제공하는 시스템을 구현하였다.

이를 통해 인공지능을 통해 분석된 기대 득점 지표는 본래 취지인 객관적인 상황 정의에 더 가까운 통계적 지표로 기능할 수 있음을 보이고, 선수들의 훈련 및 구단 운영에 기여하는 평가 요소로서의 솔루션이 될 가능성을 제기한다.

또한, 본 논문에서는 축구의 기대 득점, 즉 슈팅 상황만을 한정하여 다루었으나, 이외의 역동적인 스포츠 상황에서도 영상 및 이미지를 인공지능을 통해 처리함으로써 다양한 정보를 취합 및 가공하고 이에 대한 다양한 분석 서비스 역시 개발할 수 있을 것이다.

참고 문헌

- [1] Green, S. (2012, April 12). Assessing The Performance of Premier League Goalscorers. Retrieved from <http://www.optasprtspro.com/about/optapro-blog/posts/2012/blog-assessing-the-performance-of-premier-league-goalscorers/>.
- [2] Redmon, J., & Farhadi, A. (2018). Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767*.
- [3] Cox, D. R. (1958). The regression analysis of binary sequences. *Journal of the Royal Statistical Society: Series B (Methodological)*, 20(2), 215-232.

본 논문은 과학기술정보통신부 정보통신창의인재양성사업의 지원을 통해 수행한 ICT멘토링 프로젝트 결과물입니다.