

축구 빅데이터를 활용한 시각화와 미래의 유망선수 발굴

박기혁, 박희경, 최기준

인하대학교 통계학과

서론

야구에서는 선수의 경기 성적 및 능력을 수치화하여 데이터를 분석하여 성공한 사례들이 몇몇 있다. 축구에서도 데이터를 활용하여 인사이트를 발견하고자 한다.

선수들의 경기 이벤트를 이용하여 능력을 수치화하고 변수 축약과 시각화를 이용하여, 향후 국가대표가 될 선수를 포지션 별로 예측하는 모형을 제시한다.

Keyword : Korea Football Association, Matrix Projection, Factor Analysis, Principal Component Analysis, Logistic Regression

아이디어 및 데이터 설명

1868~2016년도 선수의 경기 이벤트 데이터

이벤트는 총 51개(파울, 오프사이드, 경고, 퇴장, 가로채기 등)로, 선수를 직관적으로 설명하기에 변수의 개수가 많으며, 변수 간의 높은 상관관계를 갖고 있기 때문에 모두 사용하기에는 부적절하다. 따라서 변수의 선택 및 차원축소가 필요하다.

IDEA 1 > 선수들의 능력 수치화 및 시각화

: 선수의 능력을 대표할 변수를 생성하기 위해, 차원 축소의 방법을 이용하여 비슷한 변수들끼리 묶고, 2차원의 평면으로 투영하여 시각화한다.

IDEA 2 > 유소년 축구 선수들의 미래 국가대표 진출 가능성 예측 모형

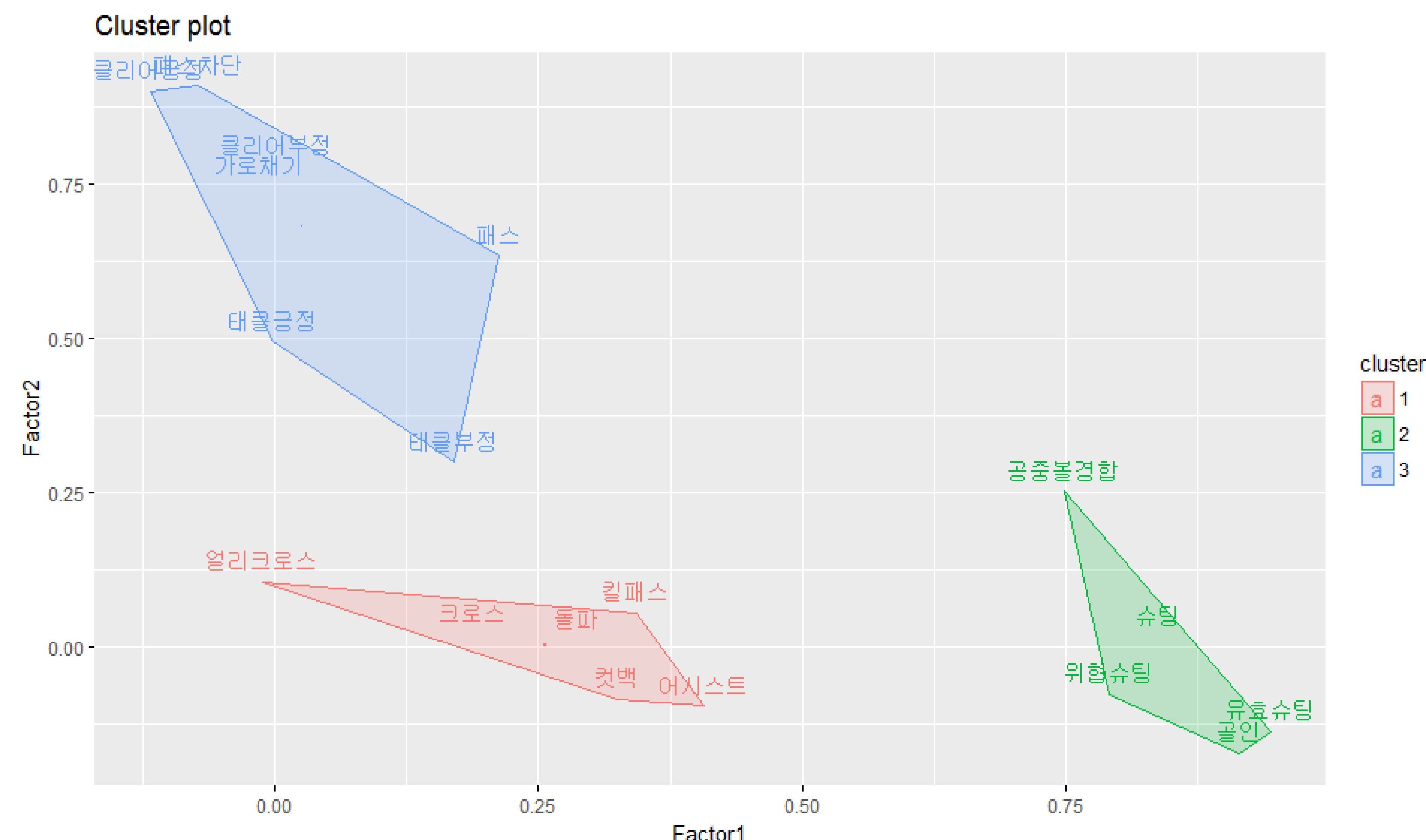
: IDEA1의 차원 축소를 통해 생성한 변수를 이용하여, 과거에 유소년 선수에서 국가대표가 된 선수의 데이터를 바탕으로 예측 모형을 세운다.

분석 과정

IDEA 1 > 선수들의 능력 수치화 및 시각화

① Factor Analysis(요인분석)를 이용한 변수 Clustering(군집화)

: 요인분석을 통한 잠재요인 2개(데이터의 약 43.8% 설명력)를 이용하여, 2차원상의 변수의 위치를 파악한다. K-means와 Hierarchical Clustering 결과 다음과 같이 변수 군집이 형성되었다.

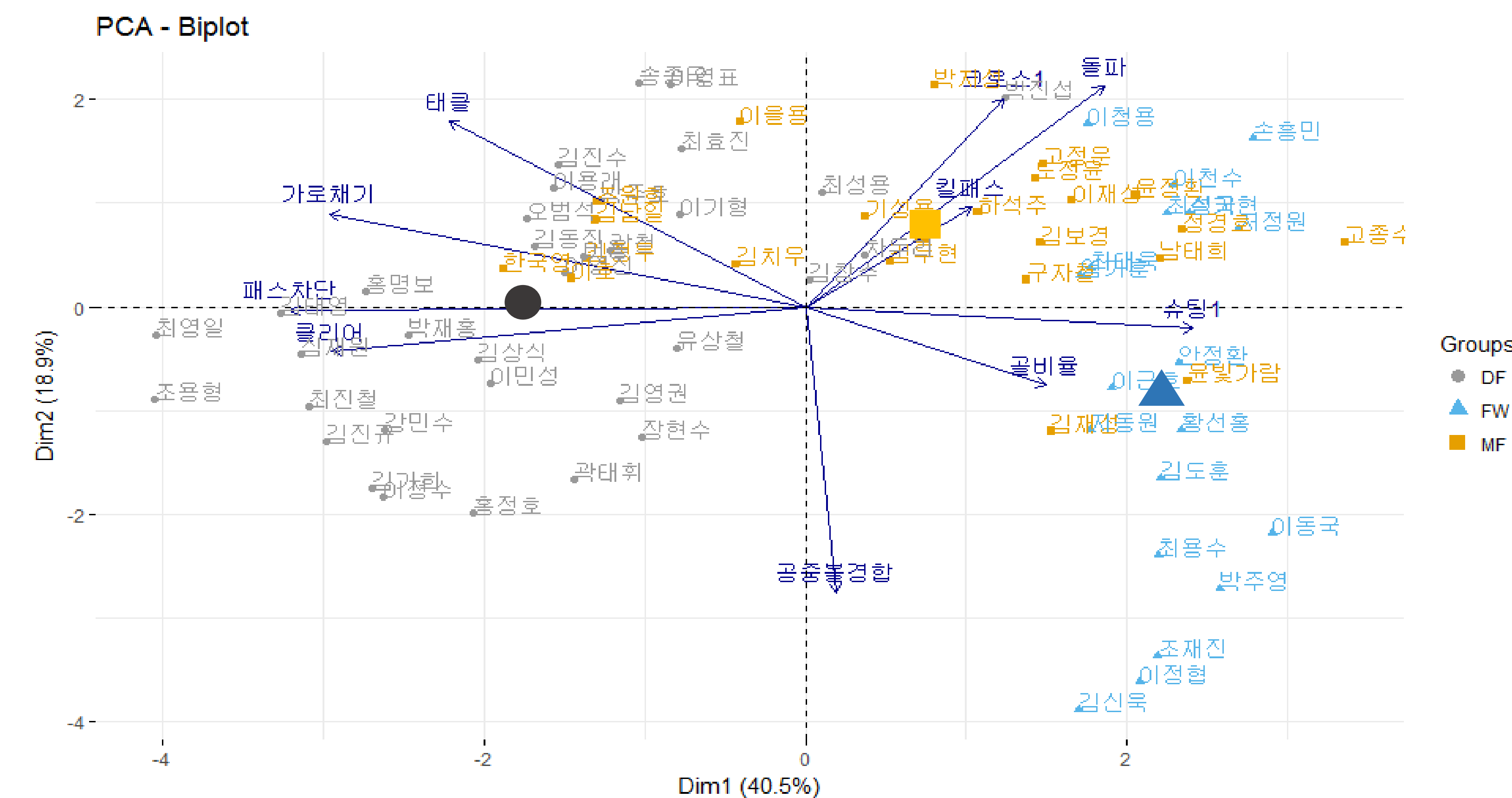


각 군집내의 변수를 이용하여, 아래와 같이 10개의 변수를 새로 생성한다.

변수명	변수 생성 방법	의미
클리어	클리어부정 + 클리어긍정	총 클리어 횟수
태클	태클긍정 + 태클부정	총 태클 횟수
크로스	크로스 + 얼리크로스 + 킥백	총 크로스 횟수
슈팅	(위험슈팅 + 유효슈팅) / 출전경기수	경기 당 슈팅 횟수
골비율	골인 / 슈팅	골 정확도
돌파	돌파 / 출전경기수	경기 당 돌파
가로채기	가로채기 / 출전경기수	경기 당 가로채기
패스차단	패스차단 / 출전경기수	경기 당 패스차단
공중볼경합	공중볼경합 / 출전경기수	경기 당 공중볼경합
킬패스	킬패스 / 출전경기수	경기 당 킬패스

② Principal Component Analysis(주성분분석)을 이용한 선수 능력 시각화

: 위에서 생성한 10개의 변수를 이용하여, 2개의 주성분(데이터의 약 59.4% 설명력)을 이용하여, 선수들의 위치를 시각화 하였다.



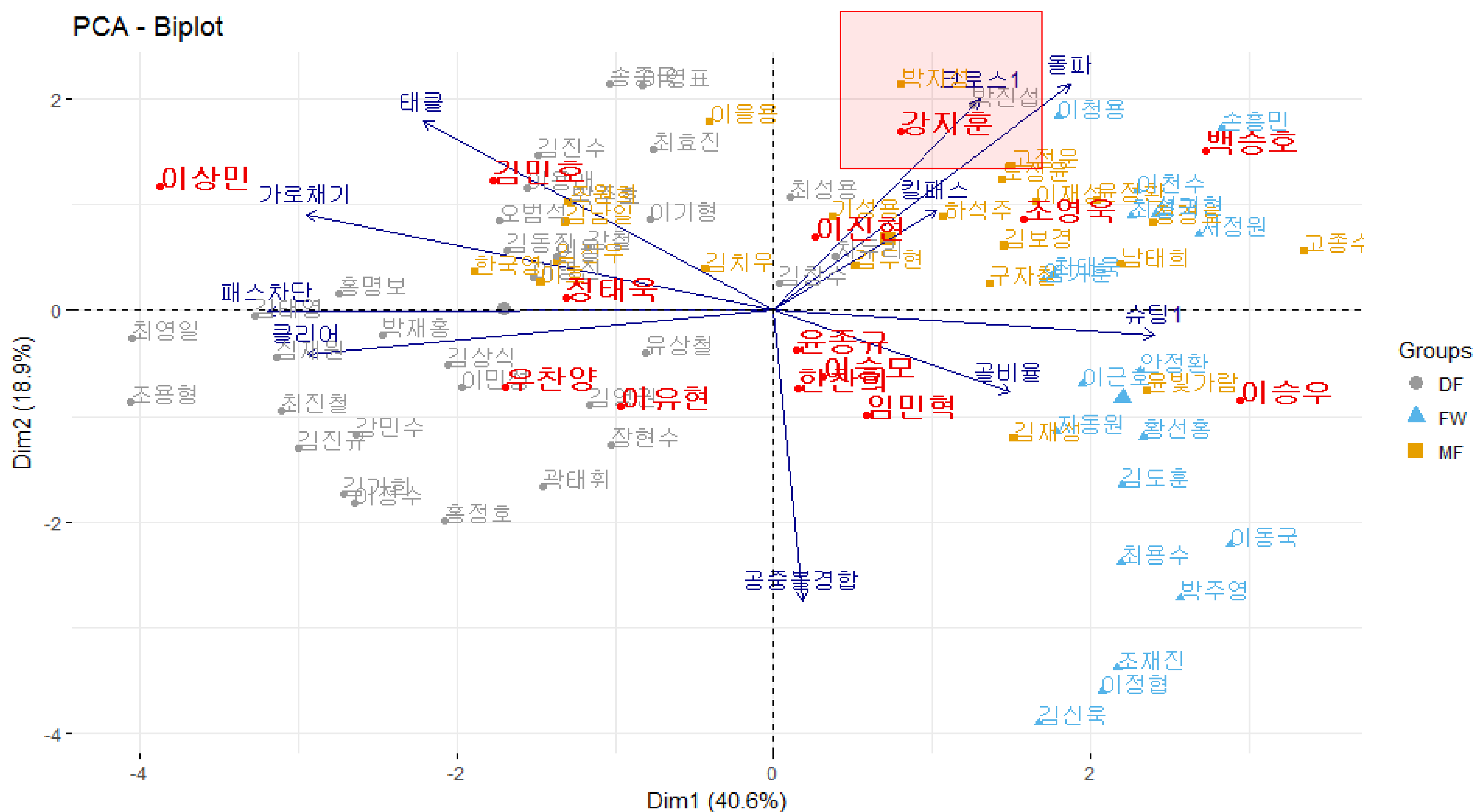
IDEA 2 > 유소년 축구 선수들의 미래 국가대표 진출 가능성 예측 모형

: 포지션 별, AIC(Akaike Information Criterion) 기준의 Backward Elimination Selection(후진제거방법)을 이용하여 Logistic Regression 예측모형을 구축한다. 포지션 별로 아래 표와 같이 변수가 선택이 되었다.

미드필더 (174명)		공격수 (97명)		수비수 (138명)	
	Estimate (Exp)		Estimate (Exp)		Estimate (Exp)
(Intercept)	-3.132 (0.044)	(Intercept)	-1.973 (0.139)	(Intercept)	-3.375 (0.034)
클리어	0.196 (1.216)	클리어	0.029 (1.029)	슈팅	2.564 (12.985)
태클	-0.053 (0.949)	크로스	0.136 (1.146)	공중볼경합	0.225 (1.291)
크로스	0.073 (1.075)	슈팅	1.437 (4.207)	돌파	0.721 (2.056)
골비율	1.27 (3.559)	돌파	-3.692 (0.025)		
가로채기	0.432 (1.541)				
공중볼경합	0.409 (1.506)				

결론

시각화를 통해, 유소년 축구 선수들이 기존의 어떤 국가대표 선수와 비슷한 성향인지를 알고, 아래와 같이 국가대표 진출 가능성에 대해 예측할 수 있다.



포지션 및 Cutoff	이름	확률	포지션 및 Cutoff	이름	확률
미드필더 Cutoff: 0.194	한찬희	0.169	수비수 Cutoff: 0.270	김승우	0.066
	백승호	0.409		이상민	0.349
	이승모	0.057		정태욱	0.267
	이진현	0.061		우찬양	0.499
	이상현	0.032		윤종규	0.344
공격수 Cutoff: 0.330	조영욱	0.406		이유현	0.51
	이승우	0.621		김민호	0.228
	강지훈	0.353		이정문	0.023

IDEA 1을 통해, 국가대표 선수와 유소년 선수의 위치를 파악한 결과는 위와 같다. 예를 들어, U-20의 강지훈 선수가 박지성 선수와 가까이 있음을 볼 수 있다. 또한, IDEA 2에서 강지훈 선수의 국가대표 진출 가능성의 확률은 0.353으로, 공격수의 Cut-off 0.330 이상으로 국가대표 진출이 가능하다고 판단이 된다. 이를 통해, 제 2의 박지성으로서 강지훈 선수의 활약을 기대해볼 만하다.