



남현수, 송현서, 이상기, 장중백

# *introducing* **ABOUT OUR TEAM**

LUX AI2 에서 '함'씨 아저씨에게 벽을 느끼고 도망

예비군 동원 훈련으로 도망간 현서님 미워



**Doraemon**  
master

# PROJECT

(batter|pitcher)2vec:  
Statistic-Free Talent Modeling With

Neural Player Embeddings(KBO ver.)

참고논문: <https://github.com/airalcorn2/batter-pitcher-2vec/blob/master/README.md>



data set

# 네이버 스포츠(야구)

네이버 스포츠 야구(중계)

<https://www.koreabaseball.com/Record/Player/HitterBasic/Basic1.aspx>

스탯티즈

<http://www.statiz.co.kr/stat.php?lr=5>

KBO 기록실

<https://www.koreabaseball.com/Record/Player/HitterBasic/Basic1.aspx>

4 타격

3 파울

2 볼

1 볼

147Km/h 직구

148Km/h 직구 | 2 - 1

147Km/h 직구 | 2 - 0

147Km/h 직구 | 1 - 0

노수광

7번타자 · 타율 0.211

타석 2 | 타수 2 | 안타 0 | 득점 0 | 타점 0 | 홈런 0 | 볼넷 0 | 피삼진 1

▶ 타석영상

노수광 : 삼진 아웃

4 스트라이크

137Km/h 슬라이더

3 볼

138Km/h 포크 | 1 - 2

2 스트라이크

148Km/h 직구 | 0 - 2

1 파울

146Km/h 직구 | 0 - 1

장진혁

6번타자 · 타율 0.167

타석 4 | 타수 4 | 안타 0 | 득점 0 | 타점 0 | 홈런 0 | 볼넷 0 | 피삼진 1

▶ 타석영상

기본

확장

가치

클러치

타석

타구1

타구2

파워

팀배팅1

팀배팅2

도루

주루

구종가치

구종구사

순	이름	팀	정렬	WAR*	G	타석	타수	득점	안타	2타	3타	홈런	루타	타점	도루	도실	볼넷	사구	고4	삼진	병살	희타	희비	비율				
																								타율	출루	장타	OPS	wOBA
1	노시환	23 한 3B	2.37	36	162	143	25	47	11	0	8	82	21	2	0	18	1	0	31	3	0	0	.329	.407	.573	.981	.451	
2	박동원	23 L C	2.08	36	140	114	18	31	5	0	10	66	25	0	0	22	1	0	23	1	1	2	.272	.389	.579	.967	.433	
3	김혜성	23 키 2B	1.93	37	166	149	28	50	8	2	1	65	15	12	0	14	1	1	22	1	0	2	.336	.392	.436	.828	.389	
4	최정	23 S 3B	1.49	37	164	136	30	39	8	0	6	65	27	2	1	17	6	1	25	3	1	4	.287	.380	.478	.858	.393	
5	러셀	23 키 SS	1.45	35	145	135	16	42	8	1	4	64	32	1	0	8	1	0	20	3	0	1	.311	.352	.474	.826	.384	
6	에레디아	23 S LF	1.44	36	154	139	16	49	8	0	4	69	28	5	5	13	1	0	21	5	1	0	.353	.412	.496	.908	.424	
7	문보경	23 L 3B	1.42	37	160	132	26	41	6	2	1	54	19	4	1	22	1	1	27	1	3	2	.311	.408	.409	.817	.381	
8	오스틴	23 L 1B	1.34	37	158	143	23	48	7	0	4	67	28	3	1	11	1	0	24	2	0	3	.336	.380	.469	.848	.396	
9	허경민	23 두 3B	1.32	34	139	121	16	37	9	0	2	52	13	3	0	12	4	2	3	3	0	2	.306	.381	.430	.811	.379	
10	양의지	23 두 C	1.32	34	136	117	13	33	7	0	4	52	20	1	0	13	5	0	18	6	0	1	.282	.375	.444	.819	.387	
순	이름	팀	정렬	WAR*	G	타석	타수	득점	안타	2타	3타	홈런	루타	타점	도루	도실	볼넷	사구	고4	삼진	병살	희타	희비	비율				
																								타율	출루	장타	OPS	wOBA
11	강민호	23 삼 C	1.29	30	124	113	15	34	6	0	5	55	21	2	0	9	2	0	20	4	0	0	.301	.363	.487	.850	.396	
12	홍창기	23 L LF	1.28	37	173	138	24	43	9	1	0	54	12	8	8	25	8	0	26	4	1	1	.312	.442	.391	.833	.405	
13	알포드	23 K LF	1.23	32	135	120	20	37	9	0	4	58	13	4	1	14	1	0	36	3	0	0	.308	.385	.483	.869	.407	
14	박성한	23 S SS	1.23	36	152	127	15	34	7	0	2	47	14	1	0	24	0	1	16	1	0	1	.268	.382	.370	.752	.362	

# WORD2VEC

- 구글에서 개발한 Word2Vec은 분포 가설(distributional hypothesis)을 가정하에 표현한 분산 표현을 따름
- 학습 과정을 쉽게 병렬화(parallelization)하여 짧은 시간 안에 양질의 단어 벡터 표상을 얻을 수 있음
- Word2Vec에는 CBoW와 Skip-Gram이라는 두 가지 학습 방법이 존재

## CBoW(Continuous Bag of Words)

- 주변에 있는 단어들로 중간에 있는 단어들을 예측하는 방법

## Skip-Gram

- 중심 단어를 보고 어떤 주변 단어가 존재하는지 예측하는 모델

# (BATTER|PITCHER) 2VEC

- word2vec 의 개념을 야구에도 접목할 수 있을까?
- KBO at-bat data from 2017, 2018, 2019, and 2020 obtained from Naver sports, stattiz, KBO official website
- Training data  
at-bats from the 2017, 2018, and 2019 seasons
- Testing data  
Exclusively previously unseen matchups  
at-bats from the 2020 season

## (batter|pitcher) 2vec: Statistic-Free Talent Modeling With Neural Player Embeddings

Michael A. Alcorn<sup>1</sup>  
Baseball  
5435

### Abstract

This paper introduces (batter|pitcher) 2vec, a neural network algorithm inspired by word2vec that learns distributed representations of Major League Baseball players. The representations are discovered through a supervised learning task that attempts to predict the outcome of an at-bat (e.g., strike out, home run) given the context of a specific batter and pitcher. The learned representations qualitatively appear to better reflect baseball intuition than traditional baseball statistics, for example, by grouping together pitchers who rely primarily on pitches with dramatic movement. Further, like word2vec, the representations possess intriguing algebraic properties, for example, capturing the fact that Bryce Harper might be considered Mike Trout's left-handed doppelgänger. Lastly, (batter|pitcher) 2vec is significantly more accurate at modeling future at-bat outcomes for previously unseen matchups than simpler approaches

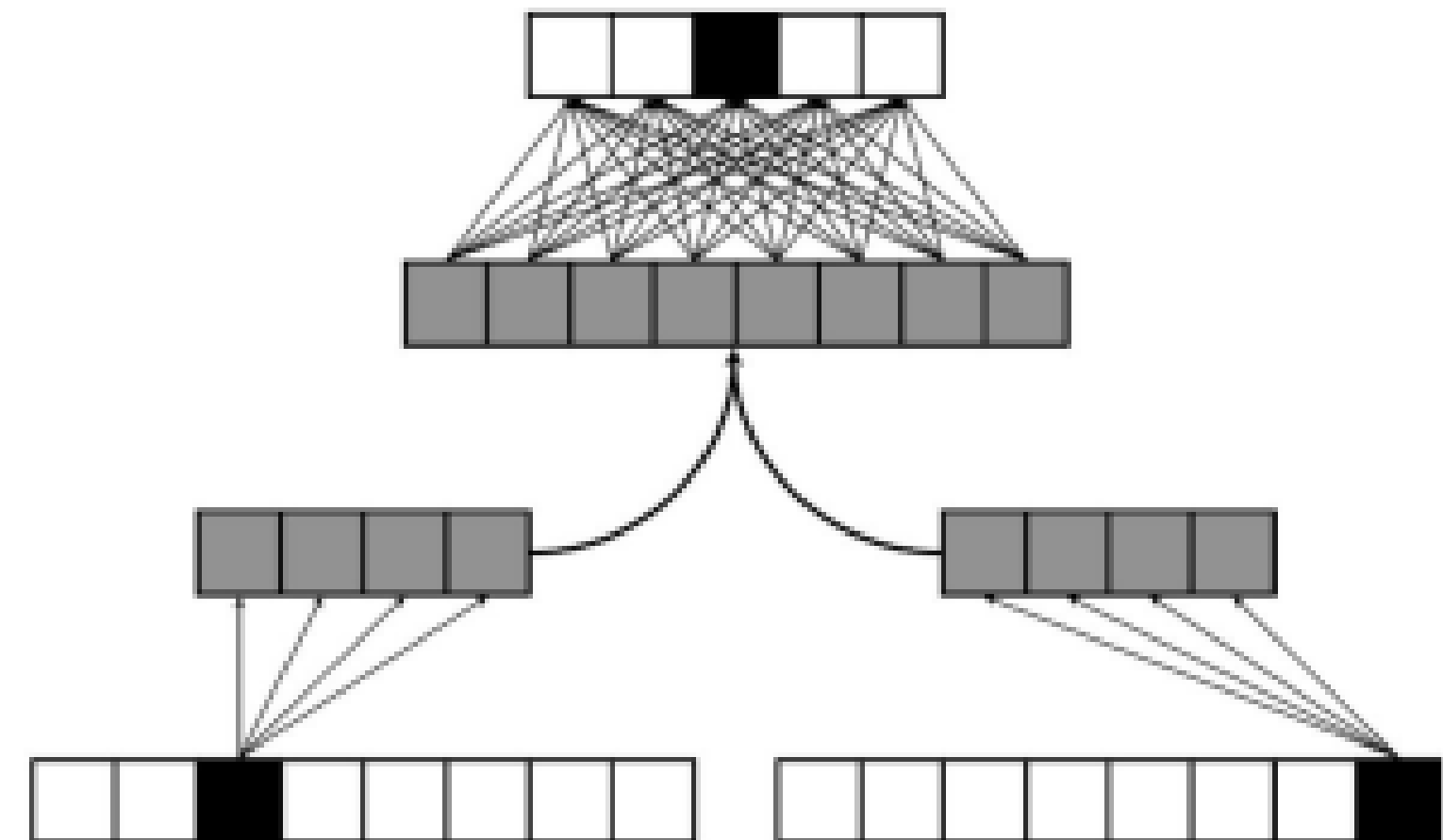
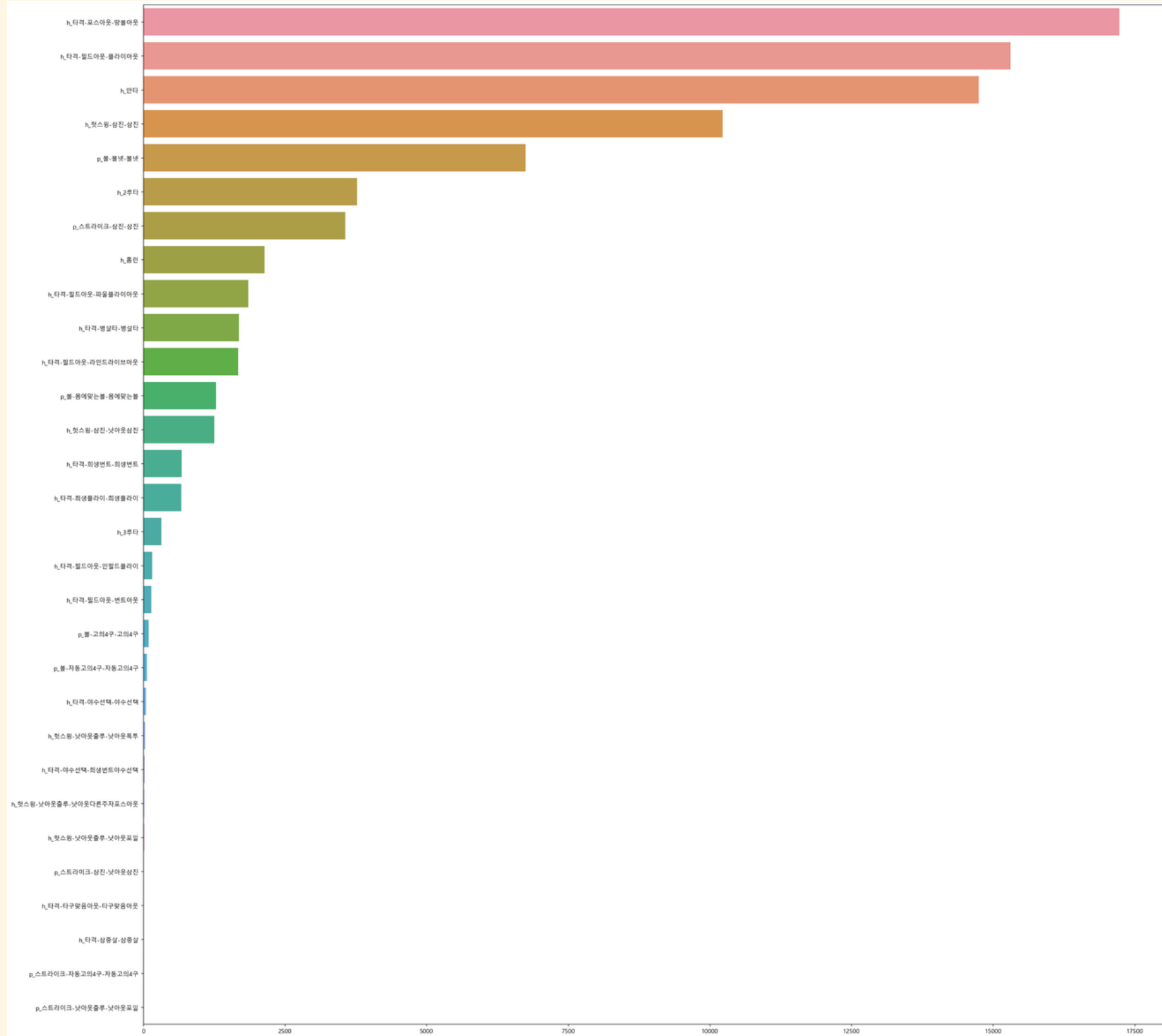


Figure 2. The (batter|pitcher) 2vec model architecture.

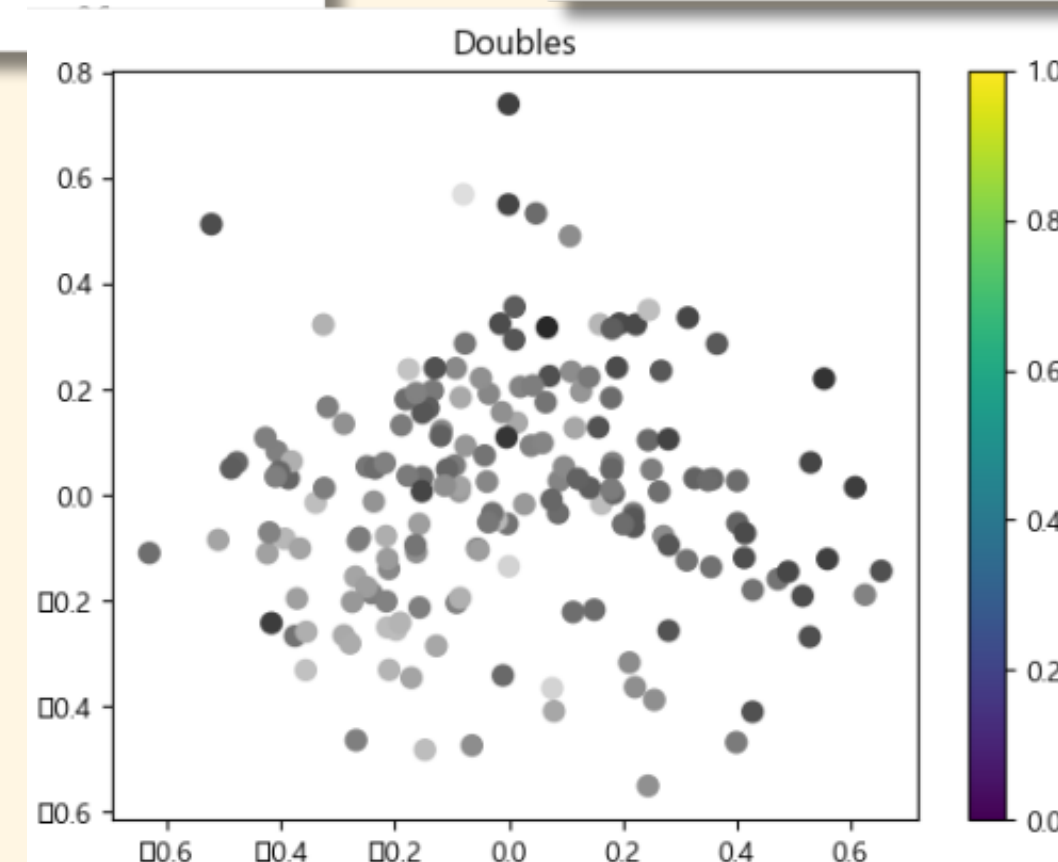
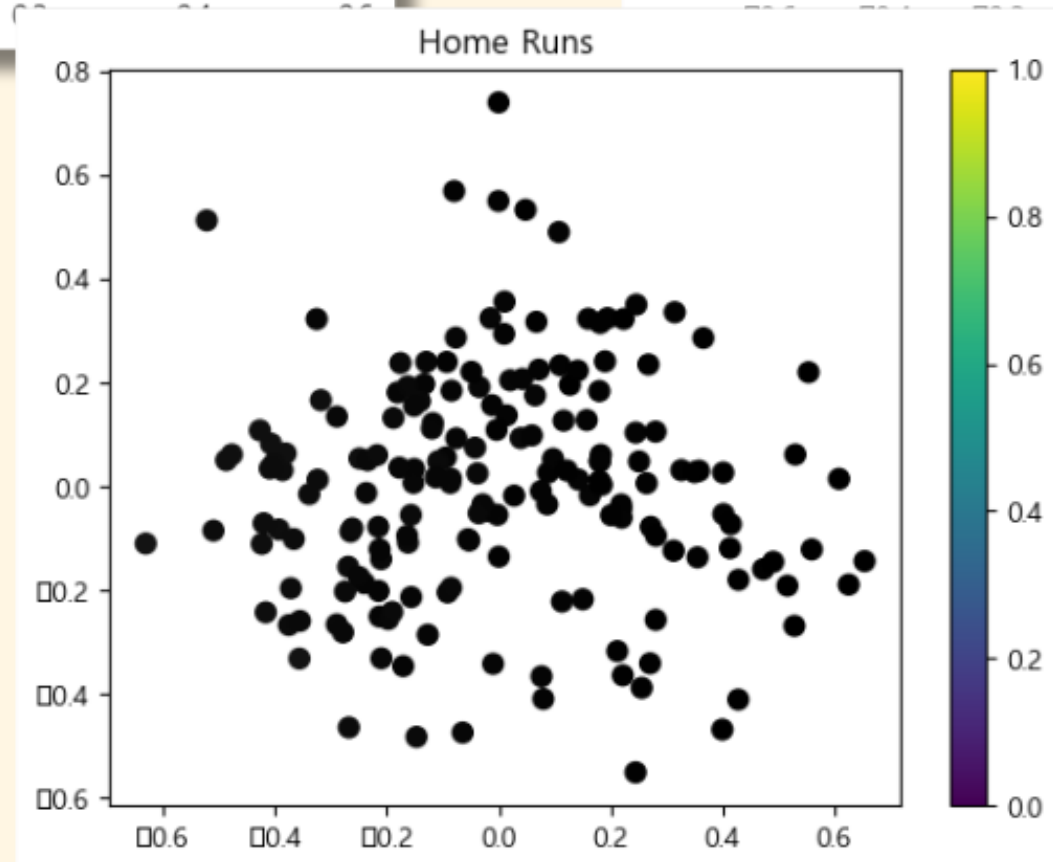
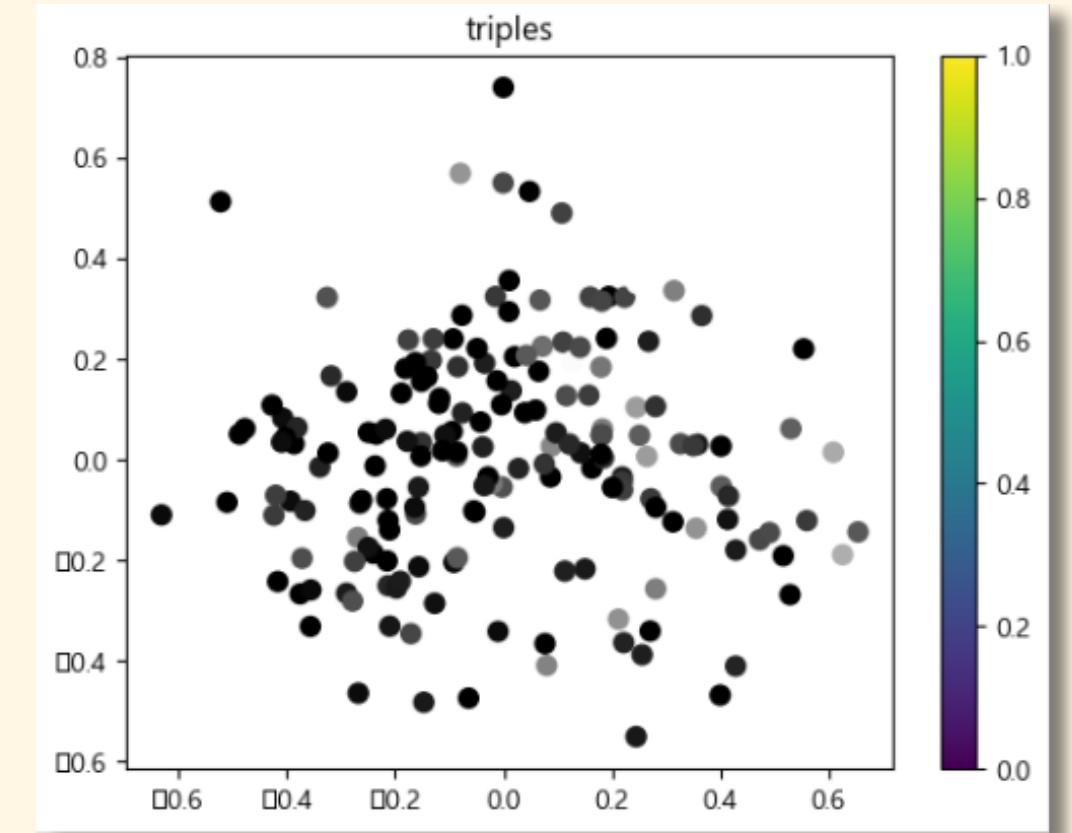
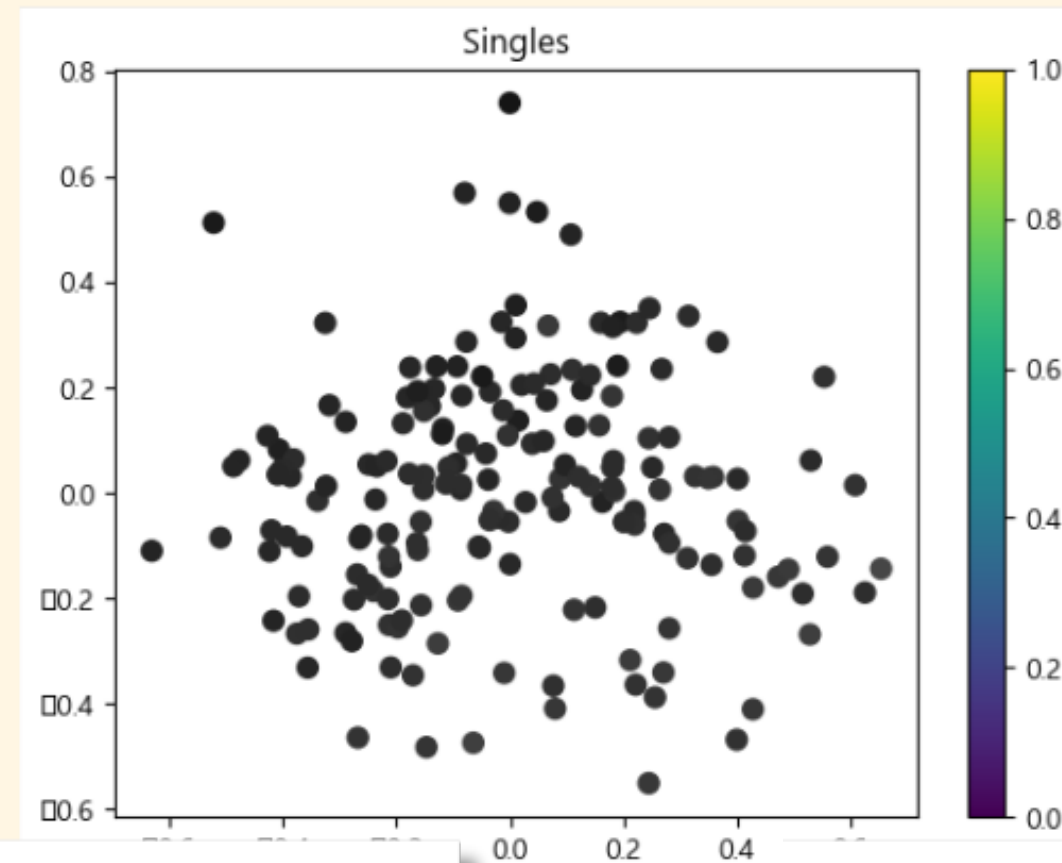
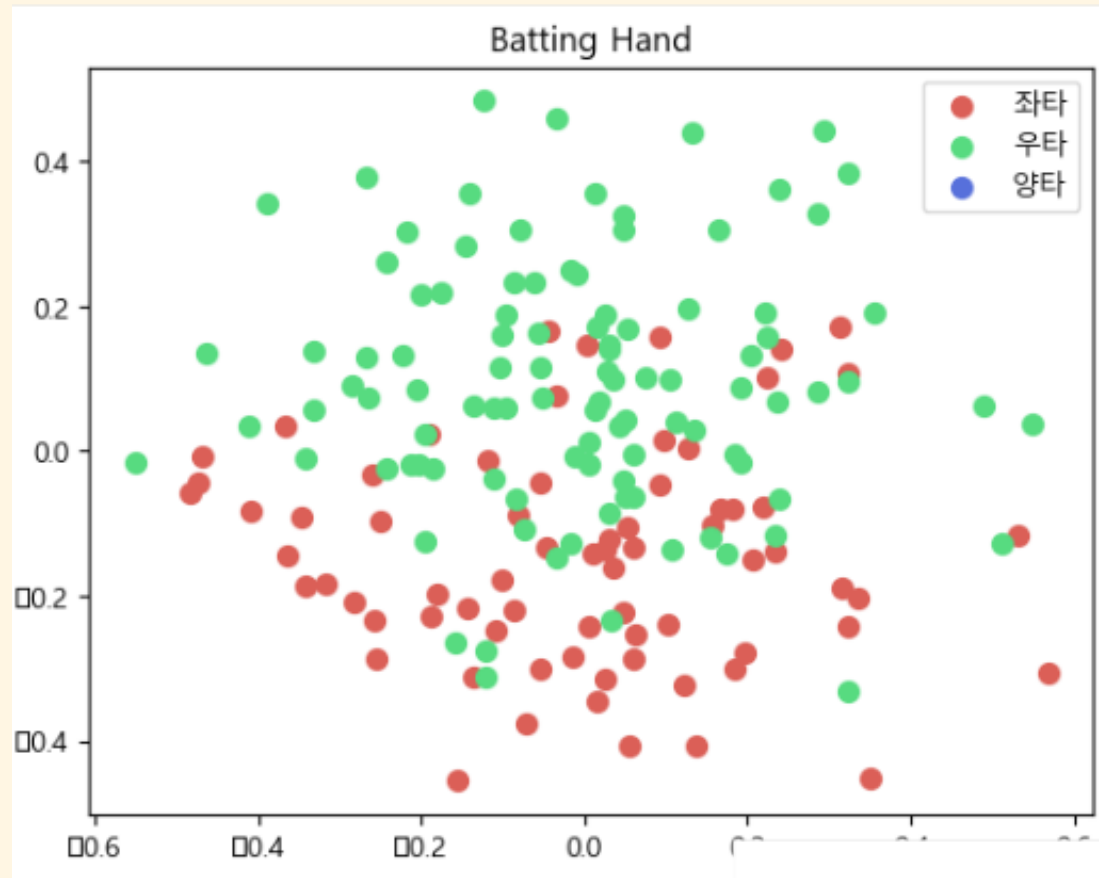
# 투수-타자 간 결과 경우의 수

- h\_타격\_포스아웃\_땅볼아웃 => h\_타격\_필드아웃\_플라이아웃=> h\_안타 => h\_헛스윙\_삼진\_삼진 순으로 빈도가 높았음.





# VISUALIZE WITH PCA





# PLAYER ALGEBRA

```
이대호
[0.5725308 0.4286211 0.4076179 0.4366924 0.7238471 0.8003677
 0.49438187 0.30808175 0.6728331 ]
[( '전준우', 0.012394952241235213), ( '정익윤', 0.012998846421735433), ( '김주찬', 0.02171221362131015), ( '김민성', 0.02176411566288794), ( '최철우', 0.022816440004657434)]

정성훈
[0.5057077 0.6694295 0.51036227 0.64732456 0.40673319 0.6485194
 0.22780643 0.5093646 0.84422827]
[( '김강민', 0.010445602988768288), ( '허정협', 0.025906983483344193), ( '최항', 0.026018009082506177), ( '베나디나', 0.027168236650979005), ( '노진혁', 0.029869302769435002)]

유희관
[0.68664974 0.37586343 0.7024373 0.5208747 0.57286847 0.48662236
 0.45458636 0.24366482 0.7895044 ]
[( '정대현', 0.01348070276089397), ( '알칸타라', 0.017751891577177936), ( '윤성환', 0.01844829140426041), ( '임기영', 0.02289620849548757), ( '손동현', 0.025027415489401017)]

양현종
[0.73341995 0.3920536 0.60181355 0.383147 0.5253383 0.67529476
 0.49669278 0.46569815 0.20757148]
[( '김광현', 0.012667637641780027), ( '왕웨이쥔', 0.013075876619118798), ( '김민수', 0.01322075556520852), ( '이영준', 0.015785868315304707), ( '터너', 0.021041891938567647)]

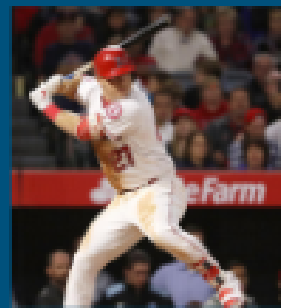
오주원
[0.63348305 0.5638122 0.8305571 0.41204298 0.59234715 0.49410278
 0.24425189 0.31611472 0.55629385]
[( '장원삼', 0.01812458291003738), ( '김선기', 0.02399005860052794), ( '이민호', 0.0280430245986385), ( '윤성환', 0.028530069360799604), ( '유희관', 0.03060657132595679)]

이대은
[0.54155236 0.63355476 0.40936753 0.42395392 0.32158655 0.7255799
 0.8077477 0.47745785 0.7173886 ]
[( '정영일', 0.007724524407240363), ( '백정현', 0.026131032552817568), ( '이민우', 0.026731372490501837), ( '원태민', 0.026846543769778797), ( '최성영', 0.02714534809971514)]
```

- 선수들 간 유사한 선수들을 찾을 수 있음.

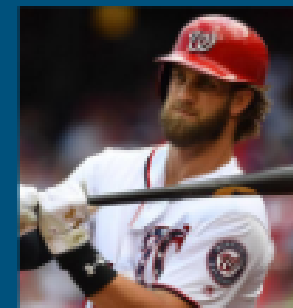
# PLAYER ALGEBRA

(batter|pitcher) 2vec



[SI.com](http://SI.com)

$$-R + L =$$



[NBCSports.com](http://NBCSports.com)

- 우타인 타자에게 우타의 변수를 제외하고 좌타의 변수를 넣어준다면 어떤 결과가 나올 것인가?

서건창

[('서건창', 0.010609603331773632), ('김현곤', 0.014307545880486328), ('정근우', 0.015603605159668299), ('오선진', 0.030821991232885537),

박용택

[('박용택', 0.01085201078237985), ('최현우', 0.011859733364109393), ('이재원', 0.017504482943135335), ('이택근', 0.021621626801793803),

**THANKS  
FOR WATCHING**