

Action Valuation in Sports: A Survey

Supplementary Material

A. Extended data analysis

In Table S1, we extend the datasets table from the main paper by providing additional details on the number of games, number of distinct actions, and variability in terms of included competitions, teams, and players. As shown, StatsBomb¹ [D1] with ED is the largest public football dataset, covering up to 3433 games and 34 different action types. Additionally, it offers large variability, spanning multiple competitions and seasons, and thus including a diverse range of teams and players. On the other hand, when extending the dataset to OTD data, StatsBomb 360 [D2] includes only a subset of these games, specifically 394. Furthermore, these games are often released in groups, meaning that for a given competition and season, only matches from a specific team are included. This limits the dataset’s applicability. Among private football datasets, [D3, D5, D7] are among the most comprehensive, both in terms of the number of games and action types available.

In other sports, such as ice hockey, there is a publicly available dataset [D9] with high variability and a large number of games. However, it is limited to ED data, lacking information about the position of the player performing the action and containing only five different action types. More complete datasets are typically provided by SportLogiq². For basketball, while the NBA³ previously made OTD data publicly available, this is no longer the case, making it difficult to access datasets with high variability and a large number of games. For other sports, its datasets tend to be smaller, and commonly including less number of games.

¹<https://statsbomb.com/>

²<https://www.sportlogiq.com/>

³<https://www.nba.com/>

Dataset	Sport	Public	Data Type			Games	Actions	Variability
			VD	OTD	ED			
[D1] - StatsBomb [†]	Football	✓	✗	✗	✓	3433	34	●●●●●
[D2] - StatsBomb 360	Football	✓	✗	✓ ¹	✓	394	34	●●●○○
[D3] - Belgian Pro League	Football	✗	✗	✓	✓	430	-	●●●●○
[D4] - Meiji J1 League	Football	✗	✗	✓	✓	55	-	●●○○○
[D5] - STATS LLC	Football	✗	✗	✓	✓	633	3 ⁺	●●●●○
[D6] - Huddl	Football	✗	✗	✓	✓	58	-	●●●○○
[D7] - Chinese Super League	Football	✗	✗	✓	✓	237	-	●●●●○
[D8] - German Bundesliga	Football	✗	✗	✓	✓	54	-	●●●○○
[D9] - NHL PBP	Ice Hockey	✓	✗	✗	✓ [*]	9220	5	●●●●○
[D10] - SportLogiq [†]	Ice Hockey	✗	✗	✗	✓	446	43	●●●●○
[D11] - NBA [†]	Basketball	✗	✗	✓	✓	784	-	●●●●○
[D12] - Badminton World Tour	Badminton	✓	✓	✓	✓	21	9	●●○○○
[D13] - German Handball League	Handball	✗	✗	✓	✗	15	0	●○○○○
[D14] - NFL PBP	American Football	✓	✗	✗	✓ [*]	~256	-	●●●●○
[D15] - Table Tennis PBP	Table Tennis	✗	✗	✗	✓ ^{**}	152	6-11	●●●●○
[D16] - StatsPerform Rugby	Rugby	✗	✗	✗	✓	1416	-	●●●●○

Table S1. Overview of datasets used in Action Valuation, detailing the sport, public availability, data type, number of games, number of action classes, and variability in terms of included competitions, teams, and players. [†] indicates multiple dataset partitions used, ^{*} indicates event positional data precision loss (categorized into field zones), and ^{**} indicates absence of positional data in events. ¹Contains only the tracking of players within the camera’s view, without the identification of those not in possession of the ball. ⁺Represents the number of action classes used from a larger, unspecified set.