Strategy Discovery in Football Soccer for the Corner Kick Use Case

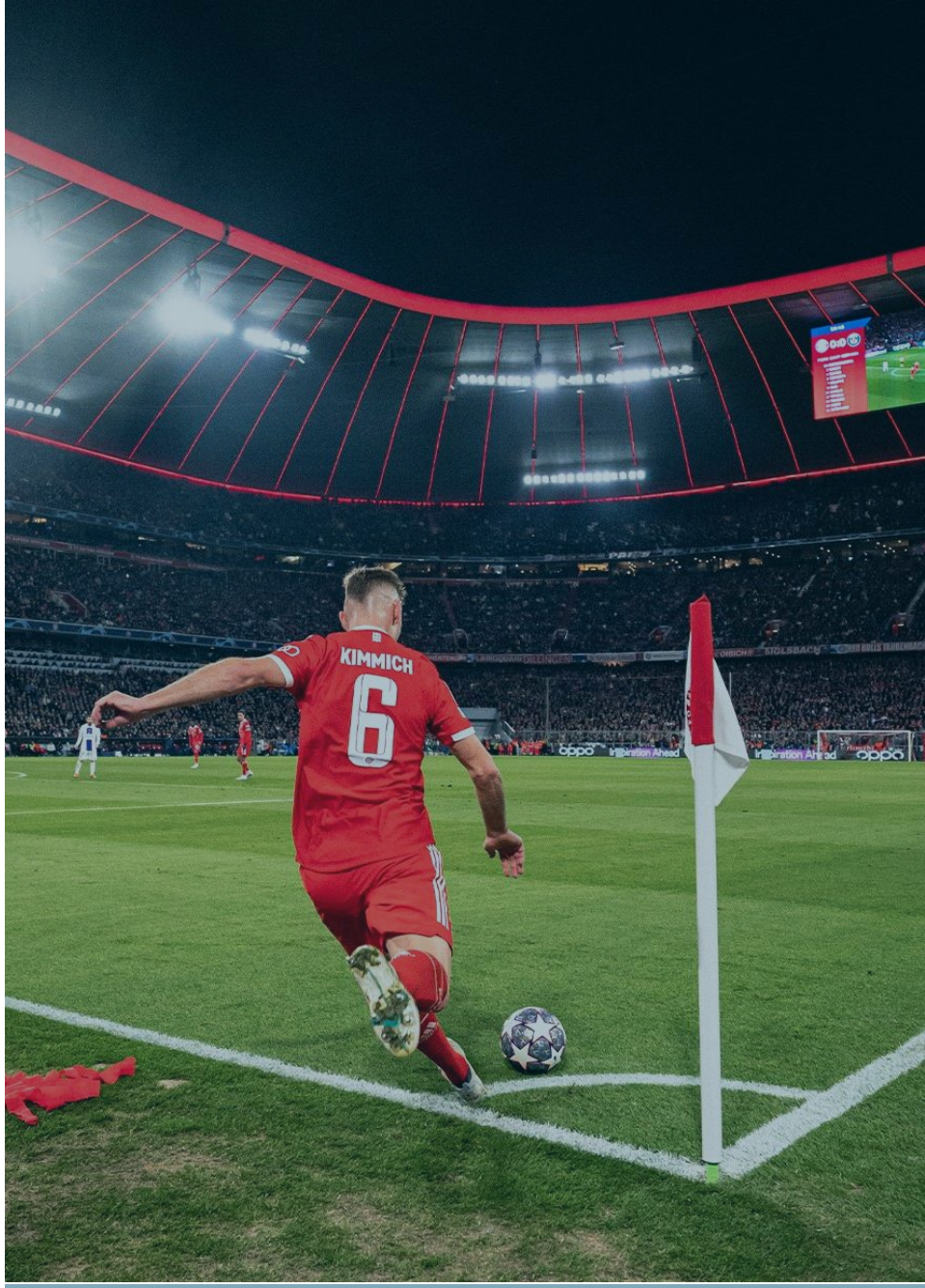
Omar Muñoz

November 2023

MSc Thesis

Outline

1. Introduction
    1. Motivation
    2. Contributions
2. Materials
3. Analysis framework
    1. Data preprocessing
    2. Discovery of tactics
    3. Conditions for tactic application
4. Conclusions and future work



I.I Motivation

- With newer sources of information there is an opportunity for further studies on data-driven decision-making.

- Approx. 2.2% of corner kicks end in goal, compared to 1.1% for open-plays.

- Limitations of current research include:
  - Disregard of the sequential context of a play.
  - Disregard of the effect of confounding variables.
  - Complex models which can be difficult to adopt by practitioners.

I.2 Contributions

Joint analysis of sequences of ball movements and their contextual conditions, to determine when and how a particular type of corner kick is more likely to succeed or not.

Summary:

1. An alternative representation of the field designed to facilitate the analysis of corner kick plays.

2. Identification and characterization of recurrent sequences of events across corner kick executions.

3. Identification and characterization of favorable and unfavorable conditions for the application of such sequences suitable for interpretation by practitioners.

2 Materials

“A public data set of spatio-temporal match events in soccer competitions.” (Pappalardo et al., 2019).

1 id	2 teamId	3 playerId	4 eventName	5 subEventName	6 tags	7 matchId	8 matchPeriod	9 eventSec	10 positions
258612104	16521	122671	'Pass'	'Simple pass'	*1x1 struct*	2057954	'1H'	1.6562	*2x1 struct*
258612106	16521	139393	'Pass'	'High pass'	*1x1 struct*	2057954	'1H'	4.4878	*2x1 struct*
258612077	14358	103668	'Duel'	'Air duel'	*2x1 struct*	2057954	'1H'	5.9374	*2x1 struct*
258612112	16521	122940	'Duel'	'Air duel'	*2x1 struct*	2057954	'1H'	6.4070	*2x1 struct*
258612110	16521	122847	'Pass'	'Simple pass'	*1x1 struct*	2057954	'1H'	8.5622	*2x1 struct*
258612113	16521	122832	'Pass'	'Simple pass'	*1x1 struct*	2057954	'1H'	10.9913	*2x1 struct*

WHO? WHAT? WHEN? WHERE?

Event data:

All on-the-ball actions along with their location on the field.

Overview:

- 1,941 matches throughout seven competitions.
- 10 different events:
  - Pass, Duel, Free Kick, Interruption, Shot, among others.
- Passes account for ~51% of the log and duels for ~27%.
- All other events account, each, for less than 8% of the log.

3.1 Data preprocessing

Log division into plays

- Identification of initial and final events.
- Focus on passing sequences.
- Removed information is stored as play metadata.
- Play labeling (success vs fail).

begin →

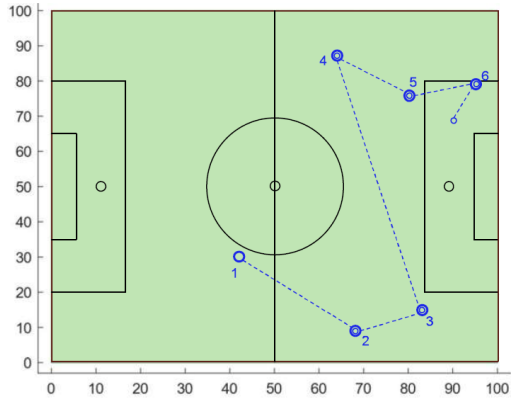
ID	Event Name	Position	Team	Sec	
1	FK (Corner)	(x_1, y_1)	A	1	Play 1 – Success
2	Pass	(x_2, y_2)	A	1.2	
3	~~Duel~~	(x_3, y_3)	A	1.5	
4	~~Shot~~	(x_4, y_4)	A	1.5	
5	Pass	(x_5, y_5)	B	5	
6	Pass	(x_6, y_6)	B	7	
7	FK (Corner)	(x_7, y_7)	A	8	Play 2 – Fail
8	Pass	(x_8, y_8)	A	8.5	
9	~~Foul~~	(x_9, y_9)	A	9.0	
…	…	…	…	…	

begin →

3.1 Data preprocessing

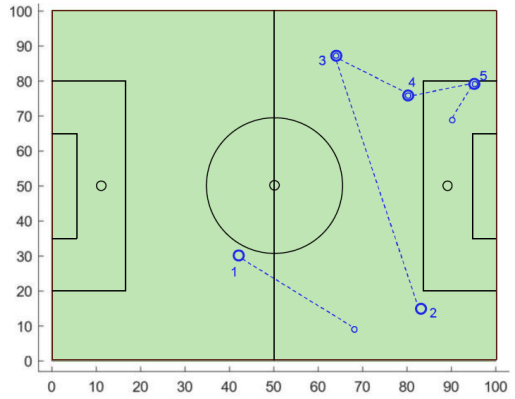
Enhance passing sequences

ID	Event Name	Initial position	Final position	Team	Sec
1	Pass	(42,30)	(68,9)	A	391
2	Pass	(68,9)	(83,15)	A	395
3	Pass	(83,15)	(64,87)	A	399
4	Pass	(64,87)	(80,76)	A	402
5	Pass	(80,76)	(95,79)	A	404
6	Pass	(95,79)	(90,69)	A	407



(a) Same location (no position mismatch).

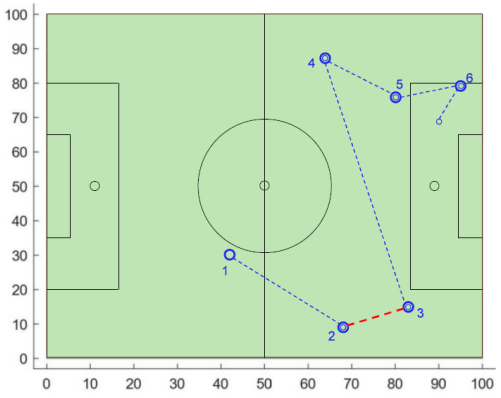
ID	Event Name	Initial position	Final position	Team	Sec
1	Pass	(42,30)	(68,9)	A	391
2	Pass	(83,15)	(64,87)	A	399
3	Pass	(64,87)	(80,76)	A	402
4	Pass	(80,76)	(95,79)	A	404
5	Pass	(95,79)	(90,69)	A	407



(b) Different location (position mismatch).

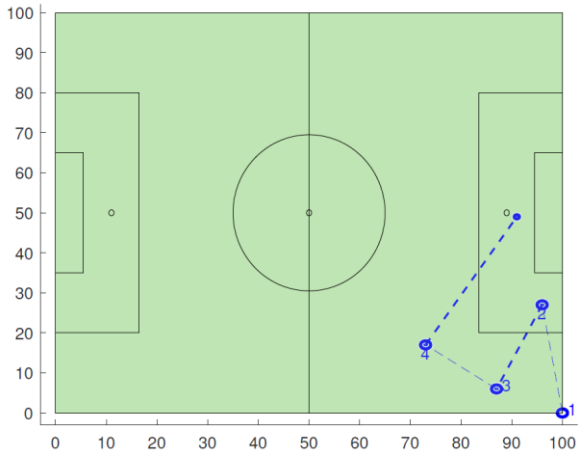
- Ball movement event allows for uninterrupted sequences of events.
- Allow us to remove initial position information due to redundancy.

ID	Event Name	Initial position	Final position	Team	Sec
1	Pass	(42,30)	(68,9)	A	391
2	Ball movement	(68,9)	(83,15)	A	395
3	Pass	(83,15)	(64,87)	A	399
4	Pass	(64,87)	(80,76)	A	402
5	Pass	(80,76)	(95,79)	A	404
6	Pass	(95,79)	(90,69)	A	407

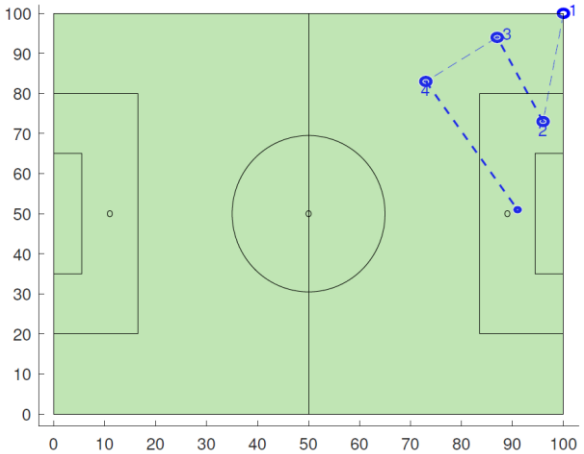


3.1 Data preprocessing

Unified frame of reference



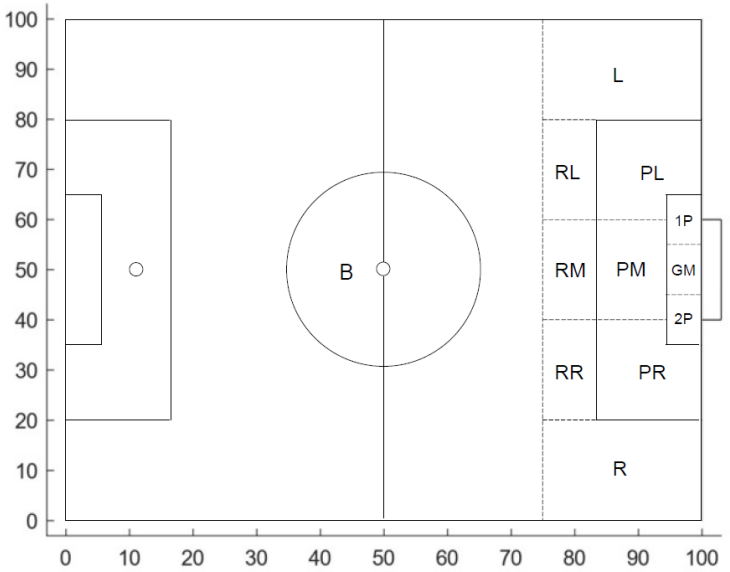
(a) Original frame of reference.



(b) New frame of reference.

- Analysis independent of the orientation.
- Increases the number of instances for our analysis.

Field discretization

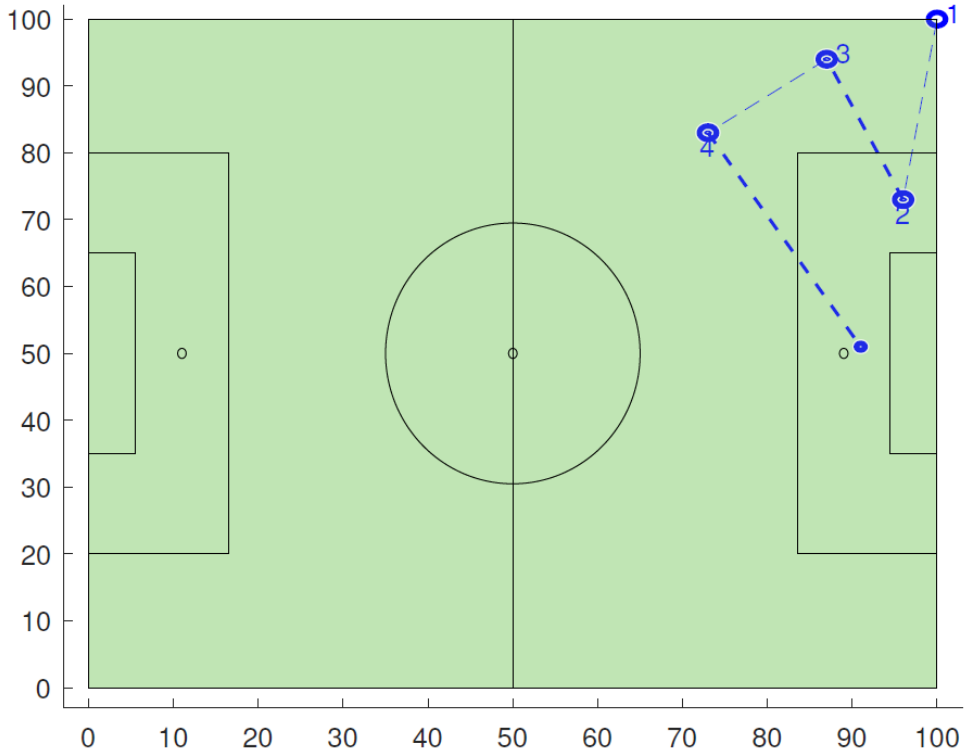


Region	Description
B	Backfield
L	Left flank
R	Right flank
RL	Rebound left
RM	Rebound middle
RR	Rebound right
PL	Penalty left
PM	Penalty middle
PR	Penalty right
1P	First post
GM	Goal middle
2P	Second post

- Express plays using domain knowledge.
- Compare similar plays.

3.1 Data preprocessing

Symbolic play creation



Example mapping:

	Play			
Event name	Pass	Pass	Pass	Ball m.
Destination region	PL	L	B	PM
Symbolic play	F	B	A	D

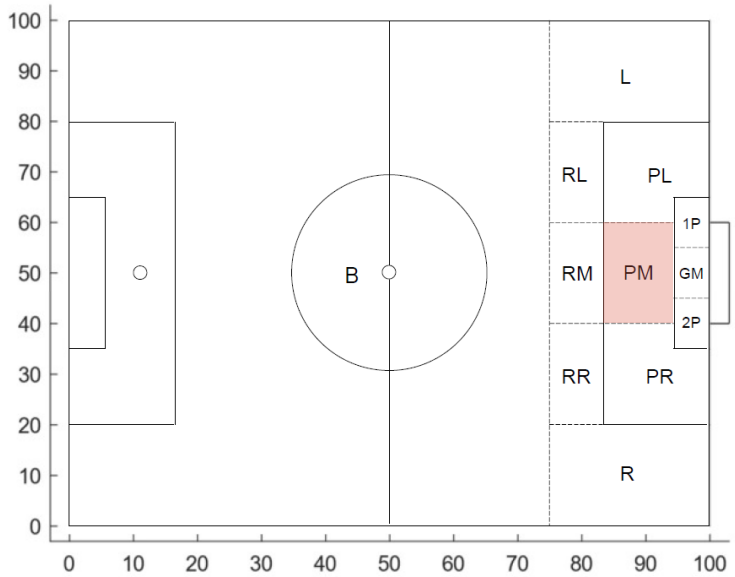
- Each (event, region) tuple is mapped to a unique symbol.

3.2 Discovery of tactics

Tactic

Recurrent sequences of events across multiple plays used by offensive teams to move the ball toward the scoring zone.

Case study selection

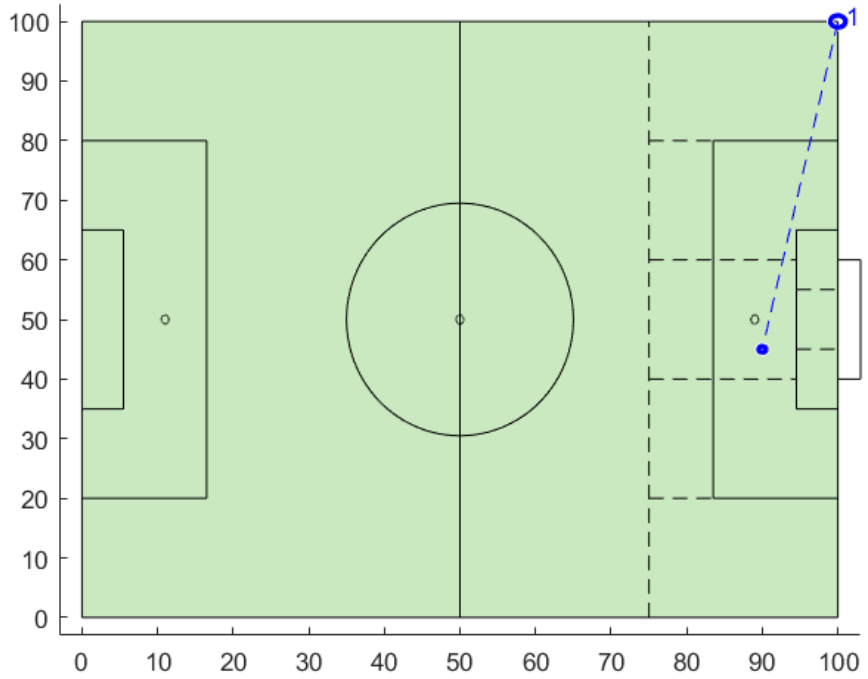


Region	Description
B	Backfield
L	Left flank
R	Right flank
RL	Rebound left
RM	Rebound middle
RR	Rebound right
PL	Penalty left
PM	Penalty middle
PR	Penalty right
1P	First post
GM	Goal middle
2P	Second post

- 6,541/17,773 corner kick plays end in region PM.

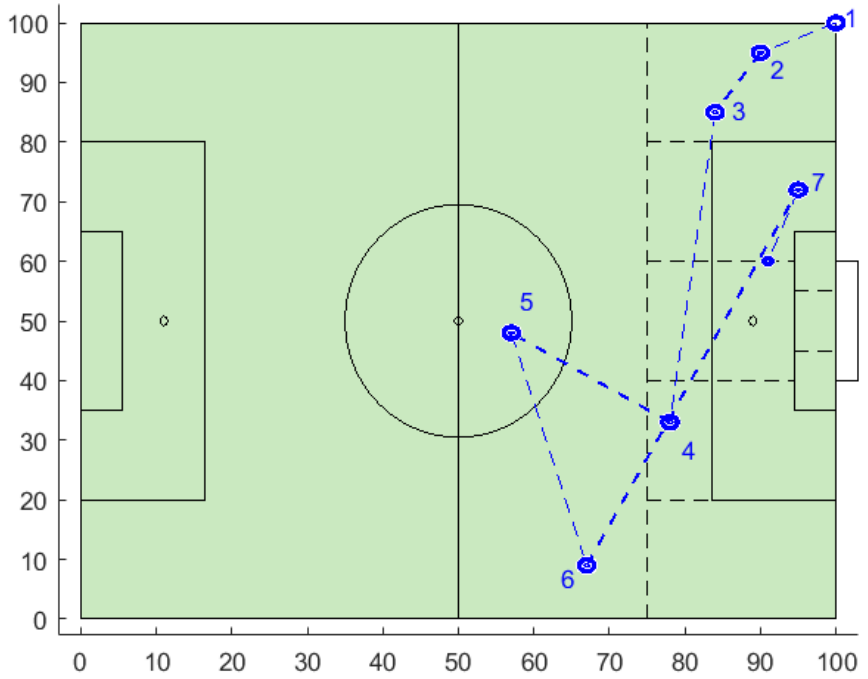
3.2 Corner kick types

Direct



- 4,974/6,541 plays.

Indirect



- 1,567/6,541 plays

3.2 Sequitur algorithm

Desirable algorithm characteristics

1. To find recurrent sequences of events in plays.
2. Establish hierarchy between recurrent sequences.
3. Express plays in terms of high-level behaviors.

Example operation

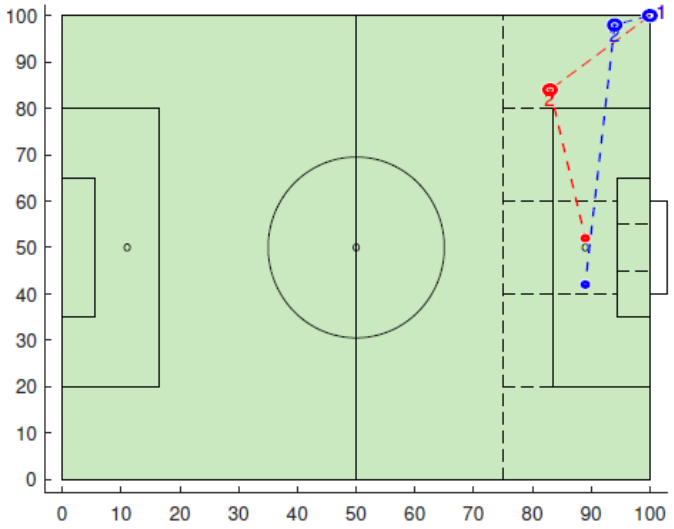
Abstract play	Symbolic play
[(Pass, B), (Pass, L), (Pass, RL)]	A B C
[(Pass, B), (Pass, L), (Pass, RL)]	A B C
[(Pass, B), (Pass, L), (Pass, RM)]	A B D
[(Pass, B), (Pass, L), (Pass, RM)]	A B D

Input: "ABC|ABC|ABD|ABD"

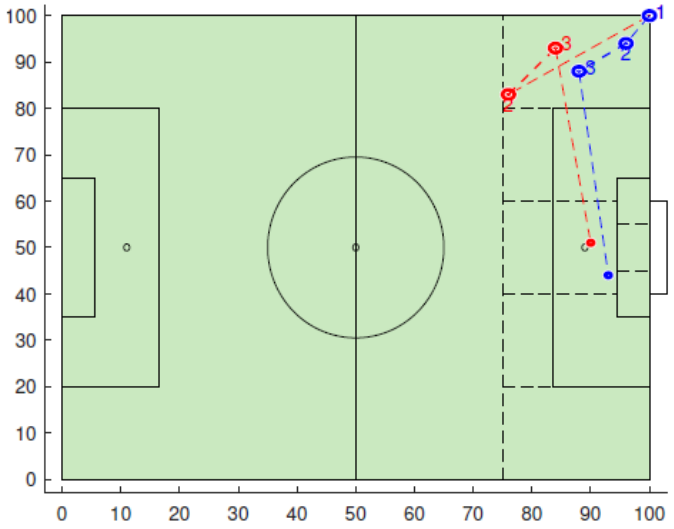
Output:

Grammar rules
R0 → R1 | R1 | R2 | R2
R1 → R3 C
R2 → R3 D
R3 → A B

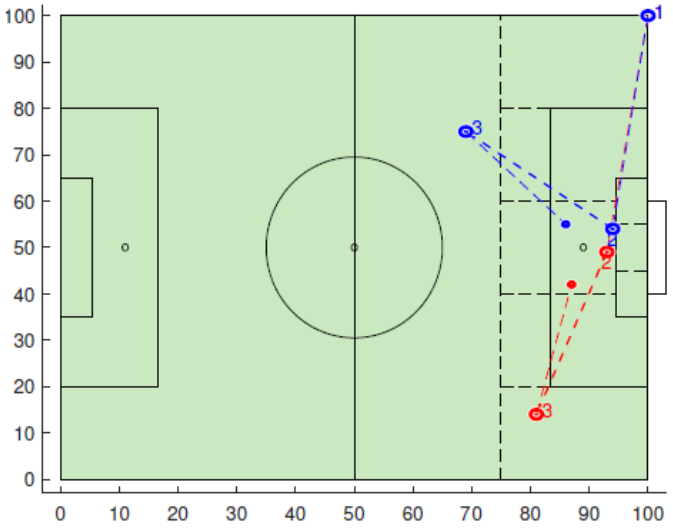
3.2 Most representative tactics (50 plays or more)



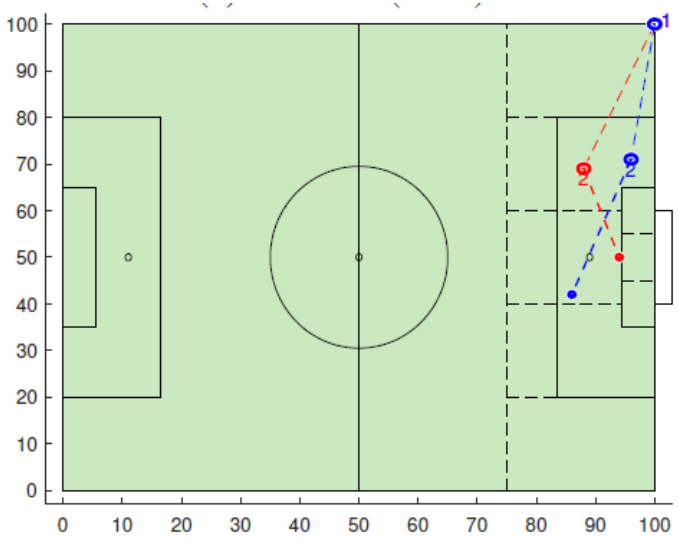
(a) Tactic 85 (short)



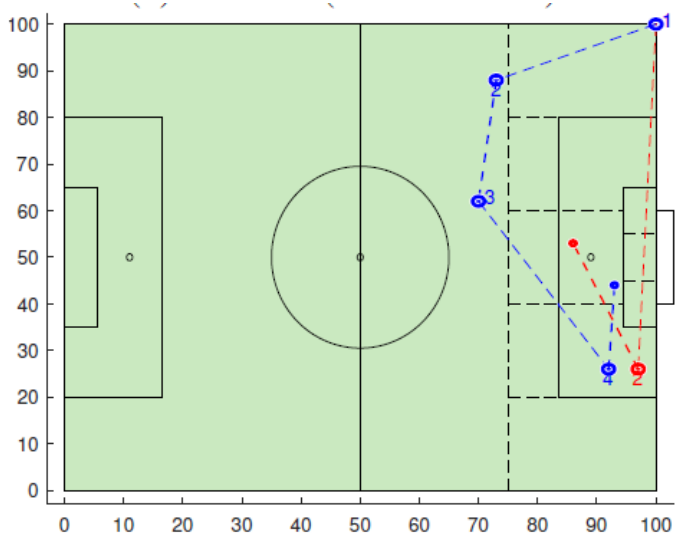
(b) Tactic 80 (short variation)



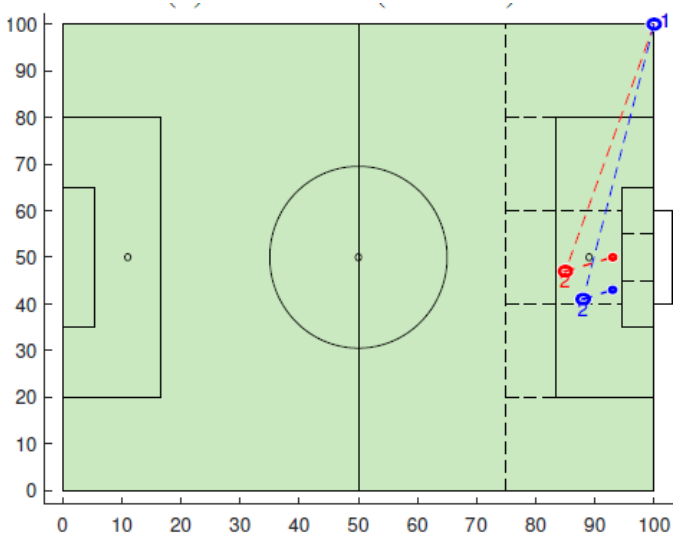
(c) Tactic 111 (rebound)



(d) Tactic 44 (near post)



(e) Tactic 141 (far post)



(f) Tactic 122 (penalty box)

3.2 Success rate

Comparisons might be biased due to different sample sizes.

Most representative:

ID	Tactic	Freq	% of indirect corners	Success rate
85	Pass left flank, Pass penalty middle	290	18.51%	0.25
80	Pass left flank, Pass left flank, Pass penalty middle	121	7.72%	0.22
111	Pass penalty middle, Ball movement, Pass penalty middle	99	6.32%	0.18
44	Pass penalty left, Pass penalty middle	97	6.13%	0.27
141	Pass penalty right, Pass penalty middle	61	3.89%	0.30
122	Pass penalty middle, Pass penalty middle	58	3.70%	0.30

We can also test the association between the tactic usage and the classes.

	Success	Failed
Tactic	a	b
~Tactic	c	d

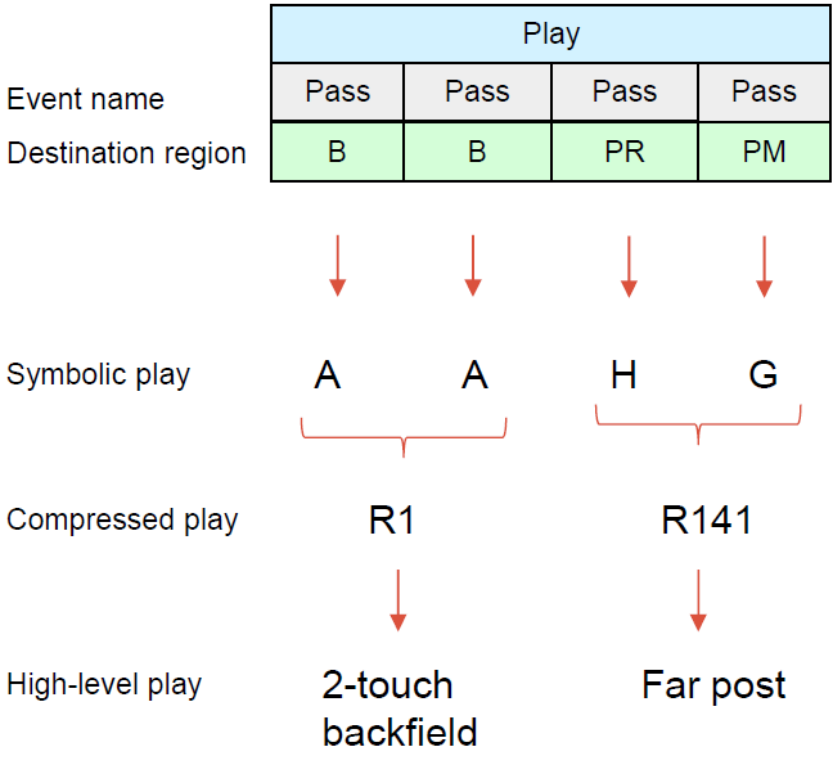
Statistically significant:

ID	Tactic	Freq	% of indirect corners	Success rate	χ^2	p-value
11	Pass backfield, Ball movement	30	1.91%	0.90	45.69	1.38E-11
30	Pass penalty left, Ball movement	49	3.13%	0.86	64.69	8.88E-16
98	Pass penalty middle, Ball movement	46	2.94%	0.80	49.14	2.38E-12
136	Pass penalty right, Ball movement	31	1.98%	0.74	24.78	6.41E-07
153	Pass first post, Ball movement	22	1.40%	0.77	20.18	7.06E-06

These share similarities with the most representative tactics but include the Ball movement event. Coaches can gain insights to exploit alternative (e.g., less common) executions.

3.2 Play compression

Example



Summary statistics

Metric	Symbolic plays	Compressed plays
Mean play length	3.17	1.23
Maximum play length	35	9
Standard deviation of play length	1.80	0.62
Mean compression factor	NA	2.61
Standard deviation of compression factor	NA	0.78

- Tactics provide a concise and informative representation of plays.
- On avg. 2.6 times shorter than the original representation.

3.3 Strategy discovery

Strategy

Contextual conditions that anticipate the success of a particular tactic.

Contextual conditions

PLAY (8)
- preparation time
- duration
- length
- number of duels
- avg. offensive height
- avg. offensive age
- offside presence
- avg. offensive market value

TEAM (7)
- avg. offensive height
- avg. defensive height
- avg. offensive age
- avg. defensive age
- avg. offensive market value
- avg. defensive market value
- defensive goalkeeper market value

USE CASE (4)
- origin corner
- preferred foot
- goalkeeper leaving line
- high corner kick

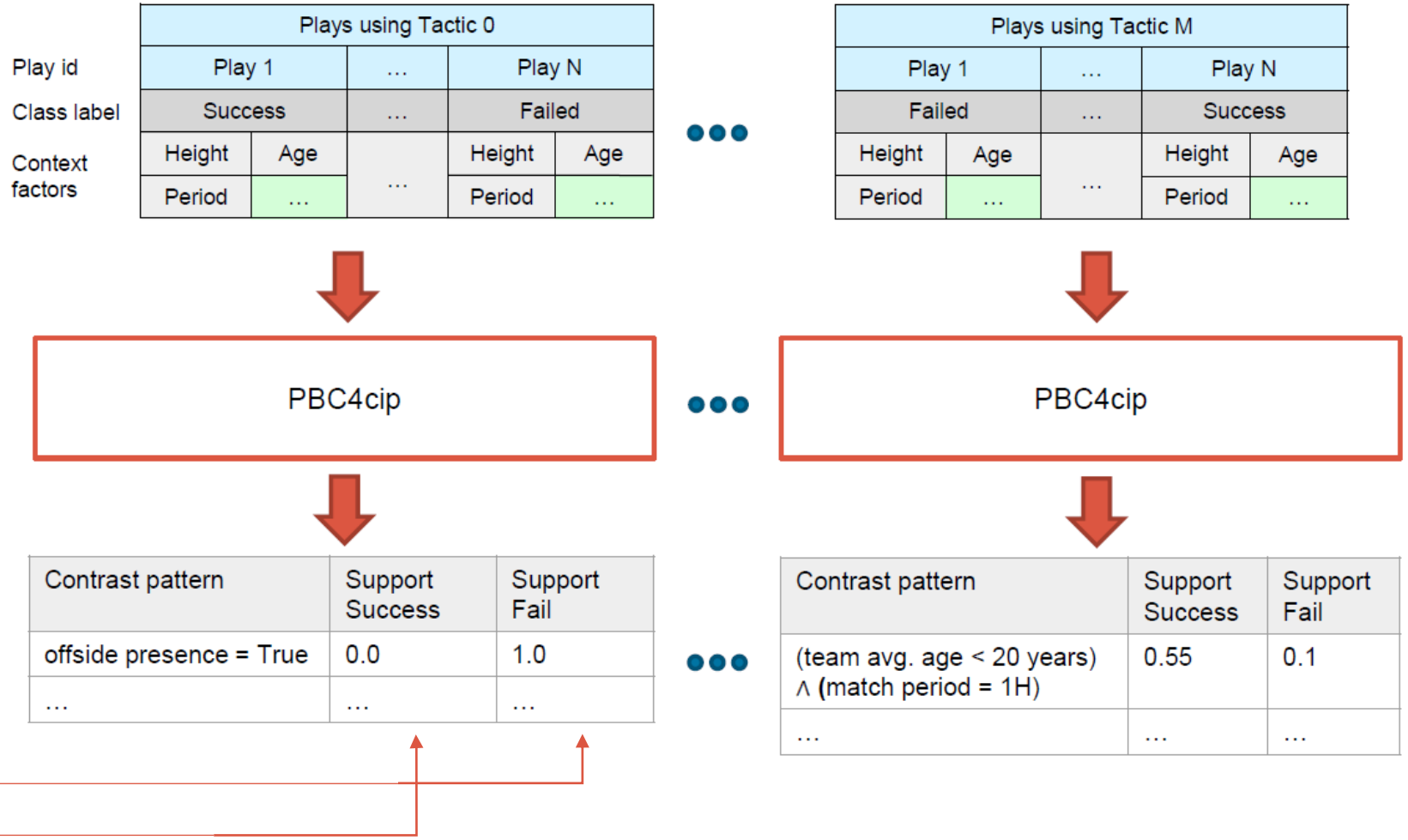
GAME (3)
- clock time
- match period
- goal difference

TOURNAMENT (2)
- progress
- advantage

- Selection based on availability, domain knowledge, and the results from prior work.

3.3 Approach

1. Contrast pattern mining per tactic.
2. Pattern filtering.
3. Pattern selection.



Quality metric

$$\mathrm{DiffSup}(X) = |\sup_{D_{s_i}}(X) - \sup_{D_{f_i}}(X)|$$

3.3 Favorable conditions for tactic application

Contrast patterns prevailed for only six out of 172 tactics.

Coaches may identify relevant contextual factors and choose a tactic that suits their needs.

E.g.:

- **Rebound** is not dependent on player height.
- **Direct** and **Short** mainly depend on player height.

Tactic name	Contrast pattern	Support		
		Success	Fail	Difference
Far post	*team avg. def. height* $\geq$ 183 cm $\wedge$ *goal difference* $\geq$ 0 $\wedge$ *play avg. off. height* $<$ 191 cm	0.83 = (15/18)	0.19 = (8/43)	0.64
Rebound	*team avg. def. market val* $\leq$ 12.15 M€ $\wedge$ *team avg. def. age* $\geq$ 27 years $\wedge$ *play duration* $<$ 10 sec	0.83 = (15/18)	0.26 = (21/81)	0.57
Near post	*play avg. off. height* $\geq$ 177 cm $\wedge$ *play avg. off. market val.* $>$ 6.15 M€ $\wedge$ *tournament progress* $\leq$ 0.85	0.85 = (22/26)	0.36 = (25/70)	0.49
Direct	*play avg. off. height* $\geq$ 186 cm $\wedge$ *play duration* $<$ 6 sec $\wedge$ *goalkeeper leaving line* = False	0.59 = (887/1515)	0.22 = (777/3459)	0.37
Short variation	*play avg. off. height* $\geq$ 177 cm $\wedge$ *preparation time* $<$ 19 sec $\wedge$ *num. duels* $\geq$ 1	0.5 = (13/26)	0.13 = (12/95)	0.37
Short	*play avg. off. height* $\geq$ 174 cm $\wedge$ *play duration* $<$ 8 sec $\wedge$ *num. duels* $\geq$ 1	0.55 = (40/73)	0.21 = (45/217)	0.34

Results suggest it might be easier to positively influence certain tactics.

Possible links between attributes and underlying reasons. E.g.:

- Height: Ability to fight for the ball.
- Age: Experience, stamina.
- Prep. time: Surprise factor.
- T. progress: Competitiveness.
- Mkt. value: Skill.
- Goal diff.: Player motivation.

3.3 Unfavorable conditions for tactic application

Results suggest a benefit in a per-tactic analysis of contextual conditions.

E.g.:

- Increased age can have either a positive or negative impact, depending on the tactic under analysis (e.g., **short** vs **rebound**).

Contrast patterns prevailed for only three out of 172 tactics.

Tactic	Contrast pattern	Support		
		Success	Fail	Difference
Near post	*play avg. off. market val.* $\leq$ 31.15 M€ $\wedge$ *goal difference* $<$ 2 $\wedge$ *def. goalkeeper market val.* $\leq$ 33.5 M€	0.42 = (11/26)	0.83 = (58/70)	0.41
Direct	(178 cm $\leq$ *play avg. off. height* $<$ 186 cm) $\wedge$ *num. duels* $<$ 2	0.12 = (178/1515)	0.51 = (1760/3459)	0.39
Short	*team avg. def. market val.* $>$ 2.15 M€ $\wedge$ *team avg. def. age* $\geq$ 25 years $\wedge$ *num. duels* $<$ 2	0.18 = (13/73)	0.52 = (113/217)	0.34

Notice that:

- While **direct** execution is the most common, teams without tall players (186 cm or more) may be interested in exploiting different tactics.

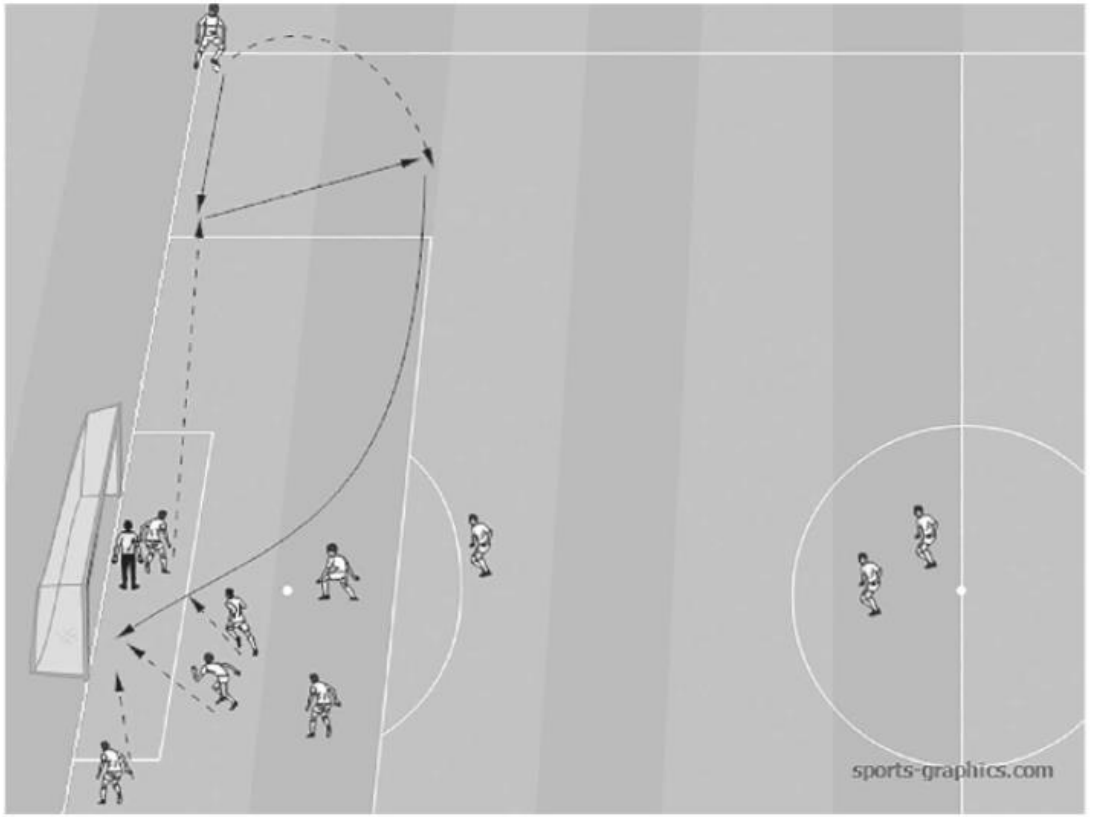
4 Conclusions

- Our representation of the field allows to express plays using a compact alphabet and terminology that is informative to the game's actors.
- Our approach to discover tactics is able to describe the step-by-step execution of known corner kick types, to gain insights into their usage, and to obtain an intuitive (compressed) representation.
- Our approach to discover strategies is driven by high-level definitions, captures both the sequential and situational context of a play, and can be easily communicated to others in natural language.
- Coaches may use our findings to support decision making or to guide further data collection.

4 Future work

- Integrate other sources of information (e.g., player tracking data) to extract sequential off-the-ball actions and additional contextual factors.
- Explore other use cases (e.g., free kicks, open-plays).
- Closely integrate football experts into the analysis.

Example of “complex” corner kick execution not captured by our approach.



ENGLUND, T. The ultimate book of soccer set pieces: Strategies for attack and defense restarts. Meyer & Meyer Sport (UK) Ltd, 2022.