



Automatically Discovering Offensive Patterns in Soccer Match Data

Vladimir Dzyuba

J.Van Haaren, S.Hannosset, J.Davis

KU Leuven, Belgium

October 24, 2015

Analysing player tracking data
requires reasoning about relations

Inductive logic programming is
a convenient tool for tackling it

But first to the most pressing issue...

Football vs. soccer

2/25

Football vs. soccer: Finger-pointing

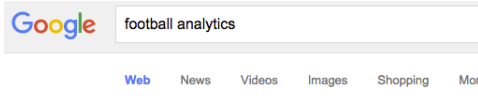
2/25

- ▶ **Jesse:** he is American

Football vs. soccer: Finger-pointing

2/25

- ▶ **Jesse:** he is American
- ▶ **Google:** "football analytics" dominated by American football



A screenshot of a Google search interface. The search bar contains the text "football analytics". Below the search bar, the "Web" tab is selected and highlighted with a blue underline. The search results show "About 27,000,000 results (0.59 seconds)". The top result is "Advanced Football Analytics" with the URL "www.advancedfootballanalytics.com/".

About 27,000,000 results (0.59 seconds)

Advanced Football Analytics

www.advancedfootballanalytics.com/

Pioneering research and analysis of NFL football. The site that makes analytics a reality.

Good Things... - Quarterbacks - Advanced Team Stats - Player Ar

Advanced Team Stats - Advanced Football Ana

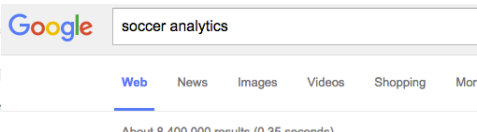
www.advancedfootballanalytics.com/index.php/.../advanced

Pioneering research and analysis of NFL football. The site that makes analytics a reality.

Fantasy Football Analytics -

fantasyfootballanalytics.net/

Using statistical analysis to help you gain an edge in your Fantasy 2015 Fantasy Football - Who has the Best Projections? - Do Stats



A screenshot of a Google search interface. The search bar contains the text "soccer analytics". Below the search bar, the "Web" tab is selected and highlighted with a blue underline. The search results show "About 8,400,000 results (0.35 seconds)". The top result is "Soccer Analytics | FiveThirtyEight" with the URL "fivethirtyeight.com/tag/soccer-analytics/".

About 8,400,000 results (0.35 seconds)

Soccer Analytics | FiveThirtyEight

fivethirtyeight.com/tag/soccer-analytics/

All posts tagged "Soccer Analytics". Chelsea manager Jose Mourinho's Barclays Premier League match between Manchester City and Ch

Soccer Analytics – Presented by Prozone | MIT

www.sloansportsconference.com/?p=9740

Soccer is undergoing a silent revolution due to the ever-increasing use of analytics in the game. Analytics are helping players become better athletes, r

Soccer Analytics: Matthew Benham's Brentford,

www.si.com/.../soccer-analytics-brentford-midtylland-matthe

May 15, 2015 - Soccer analytics revolution underway at Brentford Midtylland. Matthew Benham is sparking an analytics revolution at

Soccermetrics Research, LLC

www.soccermetrics.net/

Advanced soccer analytics for the soccer (football) industry. Empowering the

Two key drivers:

- ▶ Availability of data
- ▶ Abundance of high-profile success stories
Oakland A's (Moneyball), FC Midtjylland, etc.

Ultimate goal:

Use data analysis to improve player/team performance

Advanced data collection: Major enabling factor 4/25

High-frequency tracking data: Coordinates every 0.1s



Image: Prozone

Event streams: Manually annotated video feeds

Dataset: Games of professional Belgian team 5/25

70 games in three competitions

59 - league, 2 - cup, 9 - Europa league

2600 events per game

40+ event types, including *pass*, *run*, *receive*, *clear*...

Each event has a number of attributes

Most importantly, *time*, *coordinates*, and *player(s)*

Motivation: Tactical patterns that create shots 6/25

Goals are rare and subject to luck

Shots are more frequent and correlate to long-term success

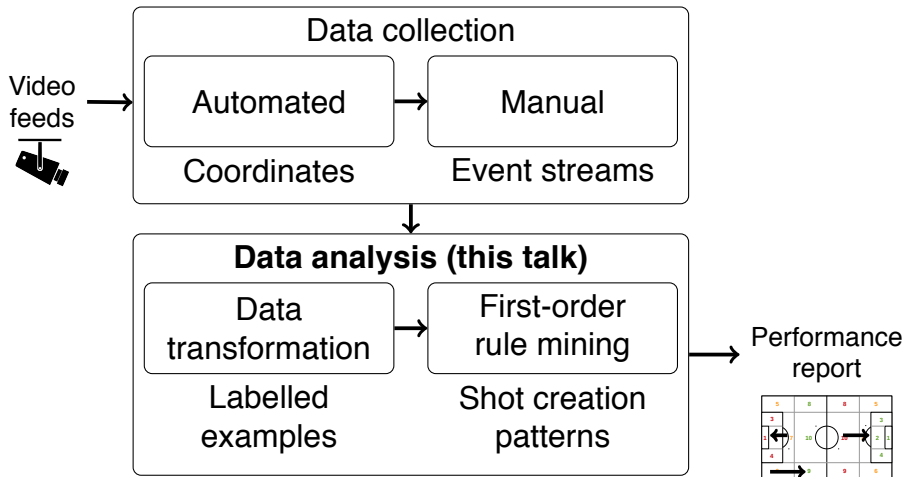
Team creating shots will eventually score

Given Game phases, labelled as *positive*,
i.e. resulted in a shot, or *negative*

Identify Features of phases
that make shots more likely

Sounds like

Rule learning, supervised pattern mining, subgroup discovery...



Outline

- 1 Data modelling & analysis
- 2 Experimental results
- 3 Conclusions

Constructing examples: Game phases

9/25

Game phases are event sequences, as determined by annotators:

Start *Throw-in, goal kick, or free kick*

End *Ball goes out of play or foul committed*



Free-kick



Pass



Pass



Pass



Cross



Pass



Shot



Pass



Clearance

















Only consider phases where *our team* is dominant

Most events in the phase are by its players

Phase is a sequence of events

















10/25

Phase label: ?

Time	Player	Team	Event	X,Y
47:35	p13	Us	 Free kick	(-1470, 2570)
47:36	p21	Them	 Reception	(-3200, 2630)
47:38	p21	Them	 Pass	(-3200, 2630)
47:40	p1	Us	 Pass	(540, 510)
47:42	p6	Us	 Pass	(-270, -1370)
47:45	p4	Us	 Pass	(-820, -2640)
47:46	p6	Us	 Reception	(-1830, -2630)
47:47	p6	Us	 Running with ball	(-1830, -2630)
47:48	p6	Us	 Running with ball	(-1830, -2630)
47:49	p6	Us	 Running with ball	(-1830, -2630)
47:51	p6	Us	 Cross	(-1830, -2630)
47:51	p26	Them	 Reception	(-4330, -770)
47:53	p6	Us	 Pass	(-4470, -2190)
47:55	p26	Them	 Clearance	(-4270, -890)
47:56	p8	Us	 Reception	(-3350, -1640)
47:57	p8	Us	 Shot not on target	(-3390, -1820)














Step 1: Only keep shots and ball movement 11/25

Phase label: ?

Time	Player	Team	Event	X,Y
47:35	p13	Us	 Free kick	(-1470, 2570)
47:36	p21	Them	 Reception	(-3200, 2630)
47:38	p21	Them	 Pass	(-3200, 2630)
47:40	p1	Us	 Pass	(540, 510)
47:42	p6	Us	 Pass	(-270, -1370)
47:45	p4	Us	 Pass	(-820, -2640)
47:46	p6	Us	 Reception	(-1830, -2630)
47:47	p6	Us	 Running with ball	(-1830, -2630)
47:48	p6	Us	 Running with ball	(-1830, -2630)
47:49	p6	Us	 Running with ball	(-1830, -2630)
47:51	p6	Us	 Cross	(-1830, -2630)
47:51	p26	Them	 Reception	(-4330, -770)
47:53	p6	Us	 Pass	(-4470, -2190)
47:55	p26	Them	 Clearance	(-4270, -890)
47:56	p8	Us	 Reception	(-3350, -1640)
47:57	p8	Us	 Shot not on target	(-3390, -1820)

Step 1: Only keep shots and ball movement 11/25














Phase label: ?

Time	Player	Team	Event	X,Y
47:35	p13	Us	 Free kick	(-1470, 2570)
47:36	p21	Them	 Reception	(-3200, 2630)
47:38	p21	Them	 Pass	(-3200, 2630)
47:40	p1	Us	 Pass	(540, 510)
47:42	p6	Us	 Pass	(-270, -1370)
47:45	p4	Us	 Pass	(-820, -2640)
47:46	p6	Us	 Reception	(-1830, -2630)
47:51	p6	Us	 Cross	(-1830, -2630)
47:51	p26	Them	 Reception	(-4330, -770)
47:53	p6	Us	 Pass	(-4470, -2190)
47:55	p26	Them	 Clearance	(-4270, -890)
47:56	p8	Us	 Reception	(-3350, -1640)
47:57	p8	Us	 Shot not on target	(-3390, -1820)














Step 2: Discard timestamps and opposing players

12/25














Phase label: ?

Time	Player	Team	Event	X,Y
47:35	p13	Us	 Free kick	(-1470, 2570)
47:36	p21	Them	 Reception	(-3200, 2630)
47:38	p21	Them	 Pass	(-3200, 2630)
47:40	p1	Us	 Pass	(540, 510)
47:42	p6	Us	 Pass	(-270, -1370)
47:45	p4	Us	 Pass	(-820, -2640)
47:46	p6	Us	 Reception	(-1830, -2630)
47:51	p6	Us	 Cross	(-1830, -2630)
47:51	p26	Them	 Reception	(-4330, -770)
47:53	p6	Us	 Pass	(-4470, -2190)
47:55	p26	Them	 Clearance	(-4270, -890)
47:56	p8	Us	 Reception	(-3350, -1640)
47:57	p8	Us	 Shot not on target	(-3390, -1820)

Phase label: ?

Player	Event	X,Y
p13	 Free kick	(-1470, 2570)
Opponent	 Reception	(-3200, 2630)
Opponent	 Pass	(-3200, 2630)
p1	 Pass	(540, 510)
p6	 Pass	(-270, -1370)
p4	 Pass	(-820, -2640)
p6	 Reception	(-1830, -2630)
p6	 Cross	(-1830, -2630)
Opponent	 Reception	(-4330, -770)
p6	 Pass	(-4470, -2190)
Opponent	 Clearance	(-4270, -890)
p8	 Reception	(-3350, -1640)
p8	 Shot not on target	(-3390, -1820)













Phase label: ?

Player	Event	X,Y
p13	 Free kick	(-1470, 2570)
Opponent	 Reception	(-3200, 2630)
Opponent	 Pass	(-3200, 2630)
p1	 Pass	(540, 510)
p6	 Pass	(-270, -1370)
p4	 Pass	(-820, -2640)
p6	 Reception	(-1830, -2630)
p6	 Cross	(-1830, -2630)
Opponent	 Reception	(-4330, -770)
p6	 Pass	(-4470, -2190)
Opponent	 Clearance	(-4270, -890)
p8	 Reception	(-3350, -1640)
p8	 Shot not on target	(-3390, -1820)

Step 3: Label phases based on shot occurrences

13/25













Phase label: ⊕

Player	Event	X,Y
p13	 Free kick	(-1470, 2570)
Opponent	 Reception	(-3200, 2630)
Opponent	 Pass	(-3200, 2630)
p1	 Pass	(540, 510)
p6	 Pass	(-270, -1370)
p4	 Pass	(-820, -2640)
p6	 Reception	(-1830, -2630)
p6	 Cross	(-1830, -2630)
Opponent	 Reception	(-4330, -770)
p6	 Pass	(-4470, -2190)
Opponent	 Clearance	(-4270, -890)
p8	 Reception	(-3350, -1640)

Step 4: Match passes with receptions

14/25










Phase label: ⊕

Player	Event	X,Y
p13	 Free kick	(-1470, 2570)
Opponent	 Reception	(-3200, 2630)
Opponent	 Pass	(-3200, 2630)
p1	 Pass	(540, 510)
p6	 Pass	(-270, -1370)
p4	 Pass	(-820, -2640)
p6	 Reception	(-1830, -2630)
p6	 Cross	(-1830, -2630)
Opponent	 Reception	(-4330, -770)
p6	 Pass	(-4470, -2190)
Opponent	 Clearance	(-4270, -890)
p8	 Reception	(-3350, -1640)

Step 4: Match passes with receptions

14/25

Phase label: ⊕

Event		From		To
 Free kick	p13	(-1470, 2570)	Opponent	(-3200, 2630)
 Pass	Opponent	(-3200, 2630)	p1	(540, 510)
 Pass	p1	(540, 510)	p6	(-270, -1370)
 Pass	p6	(-270, -1370)	p4	(-820, -2640)
 Pass	p4	(-820, -2640)	p6	(-1830, -2630)
 Cross	p6	(-1830, -2630)	Opponent	(-4330, -770)
 Pass	Opponent	(-4330, -770)	p6	(-4470, -2190)
 Pass	p6	(-4470, -2190)	Opponent	(-4270, -890)
 Pass	Opponent	(-4270, -890)	p8	(-3350, -1640)

Hard to represent as fixed-length feature vector

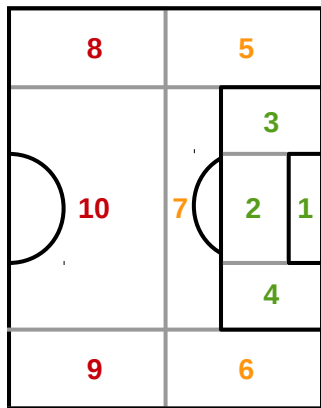
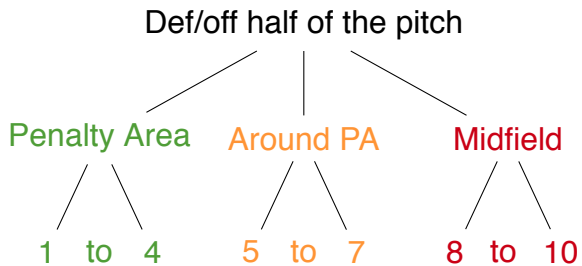
Solution: *Inductive logic programming*

Raw data is overly fine-grained:

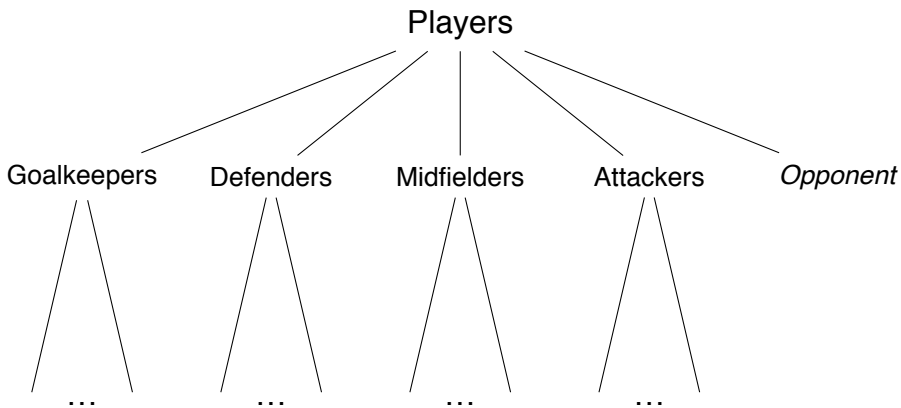
- ▶ Players are interchangeable
- ▶ Precise coordinates hardly matter

Solution: *Hierarchical background knowledge*

Zone hierarchy divides pitch into 6 zones and 20 subzones 16/25



Player hierarchy divides players by positions 17/25



- ▶ Predicates *shot/1* (target), *pass/5*, *cross/5*, and *setpiece/5* + hierarchies
- ▶ Seed-based rule search
Each example as a seed
- ▶ Guide by *m*-estimate, smoothed estimate of rule precision
- ▶ Eliminate redundant rules via post-processing

Outline

1 Data modelling & analysis

2 Experimental results

3 Conclusions

- ▶ 70 games
- ▶ 3,803 phases
 - ▶ 526 positive examples / shots (13.8%)
 - ▶ 3,277 negative examples
- ▶ 26,338 events
- ▶ 7 events per phase on average
6.5 passes, 0.3 crosses, 0.2 set pieces

Overview of results

20/25

	Setup	Hierarchy	Rules	m-est. of prec. (top 10)		Time (min.) per seed
				Maximum	Average	
/3	Spatial		276	0.7396	0.6638	0.002
		+	323	0.7396	0.7065	0.840
/3	Players		91	0.7396	0.4855	0.006
		+	257	0.7396	0.6606	5.250
/5	Combined	+				<i>Timeout</i>

Background knowledge increases runtime

21/25

526 seeds (shots)

Setup	Hierarchy	Rules	m-est. of prec. (top 10)		Time (min.) per seed
			Maximum	Average	
/3 Spatial		276	0.7396	0.6638	0.002
	+	323	0.7396	0.7065	0.840
/3 Players		91	0.7396	0.4855	0.006
	+	257	0.7396	0.6606	5.250
/5 Combined	+				Timeout

Background knowledge improves rule quality 22/25

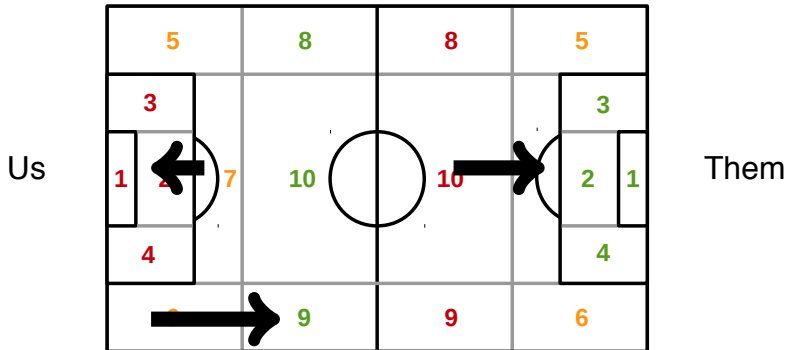
Setup	Hierarchy	Rules	m-est. of prec. (top 10)		Time (min.) per seed
			Maximum	Average	
/3 Spatial		276	0.7396	0.6638	0.002
	+	323	0.7396	0.7065	0.840
/3 Players		91	0.7396	0.4855	0.006
	+	257	0.7396	0.6606	5.250
/5 Combined	+				<i>Timeout</i>

Counter-attack via right flank and middle

23/25

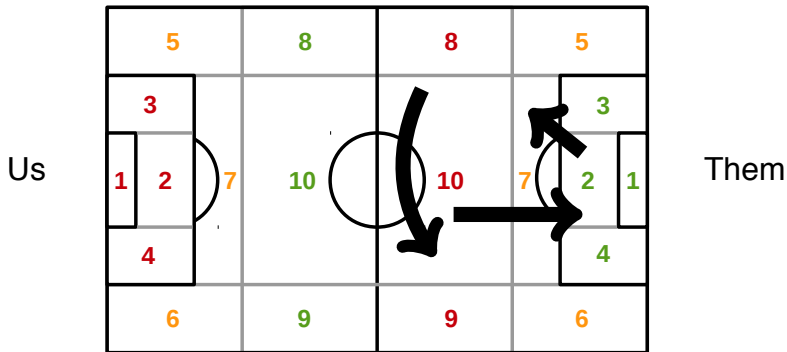
Top-scoring spatial rule

$$\begin{array}{ccc} |C| & |C^+| & m\text{-est} \\ 5 & 5 & 0.734 \end{array}$$



WRAcc gives more frequent/less pure rules 24/25

$$\begin{array}{cccc}
 |C| & |C^+| & m\text{-est} & WRAcc \\
 62 \gg 5 & 18 \gg 5 & 0.275 \ll 0.734 & 0.025
 \end{array}$$



Outline

1 Data modelling & analysis

2 Experimental results

3 Conclusions

ILP is a useful tool for modelling sports data

- ▶ Naturally relational problem
- ▶ Hard to represent with fixed-length feature vectors
- ▶ Hierarchical background knowledge

Interpretable rules engage domain experts

Room for improvement:

Leverage event order

Leverage full tracking data, e.g. relative positions

Explore rule quality measures

Automatically Discovering Offensive Patterns in Soccer Match Data

Thank you for your attention!

May I answer any questions?