Automatic Discovery of Tactics in Spatio-Temporal Soccer Match Data

Tom Decroos
Jan Van Haaren
Jesse Davis

KDD'18

Meet Bob, an analyst at Everton FC

Meet Alan, the coach at Everton FC

What Bob’s job looks like now

We are playing Manchester City next week,
I need a summary of their tactics in 3 days!

Bob manually reviews many hours of soccer video footage

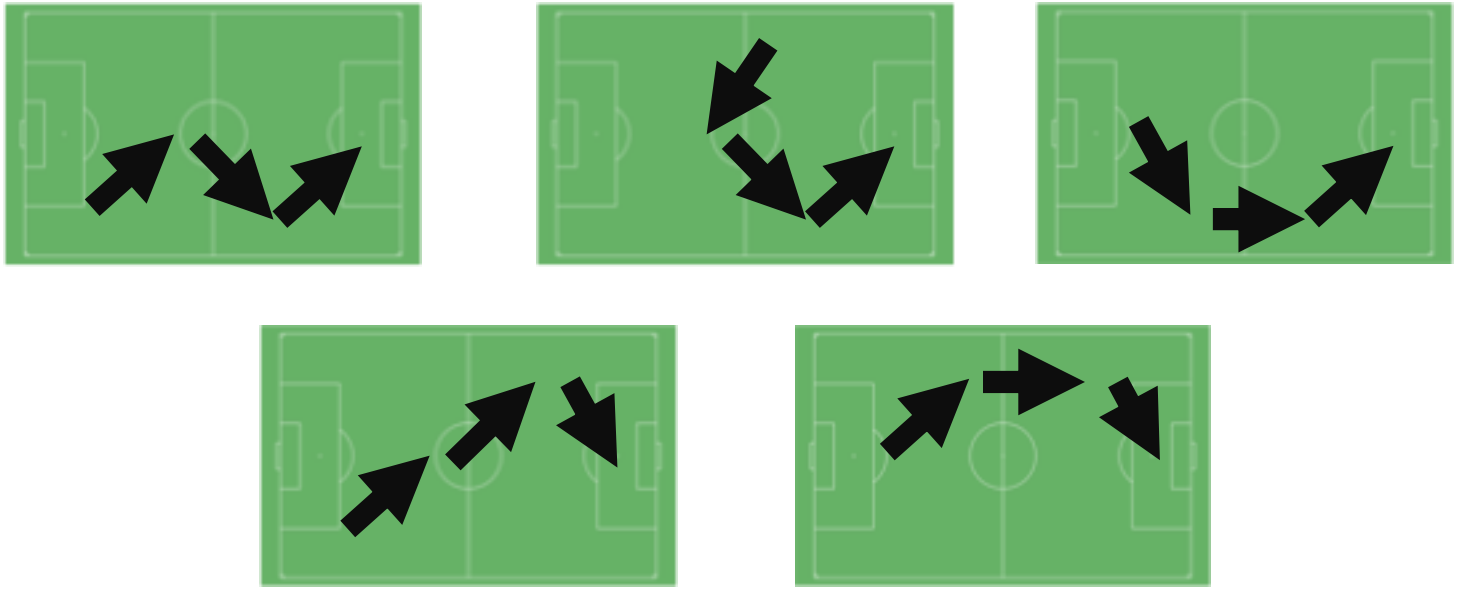
Tactics report

Thanks for you hard work.
Can you do it again for our next match?

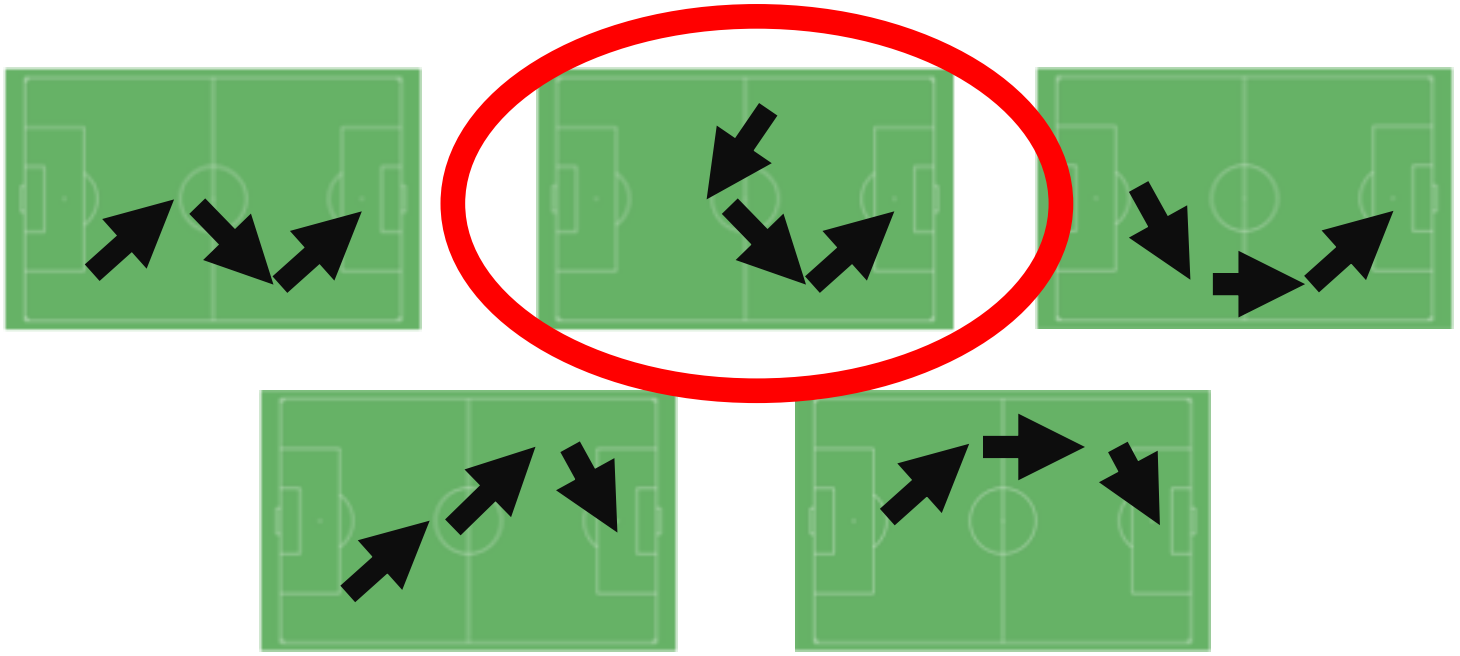
.....
Tactics report

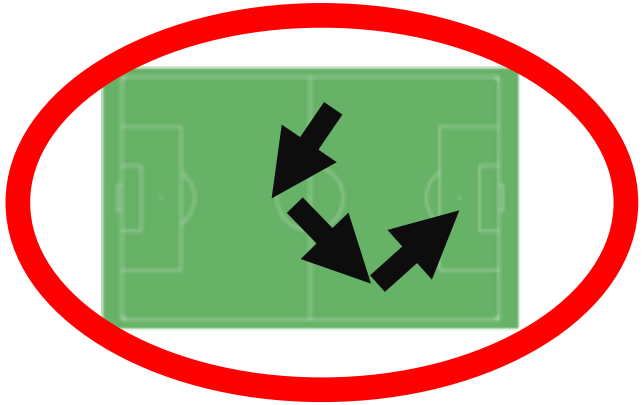
What Bob's job should look like

Computer, give me the top 5 tactics of Manchester City



Give me some examples of this tactic





Bob lets the computer automatically discover tactics

Automatic discovery of tactics in spatio-temporal soccer match data

Data

Challenges

Approach

Results

Automatic discovery of tactics in spatio-temporal soccer match data

Data

Challenges

Approach

Results

A soccer match is described by a sequence of +-1750 on-the-ball actions

A soccer match is described by a sequence of +-1750 on-the-ball actions

	Pass
Time:	12min 36sec
Location:	(32.2, 85.1)
Player:	Neil Taylor
Team:	Swansea

Automatic discovery of tactics in spatio-temporal soccer match data

Data

Challenges

Approach

Results

Tactics involve space and time

Frequent itemsets : {Pass, Goal, Pass, Dribble, Tackle}

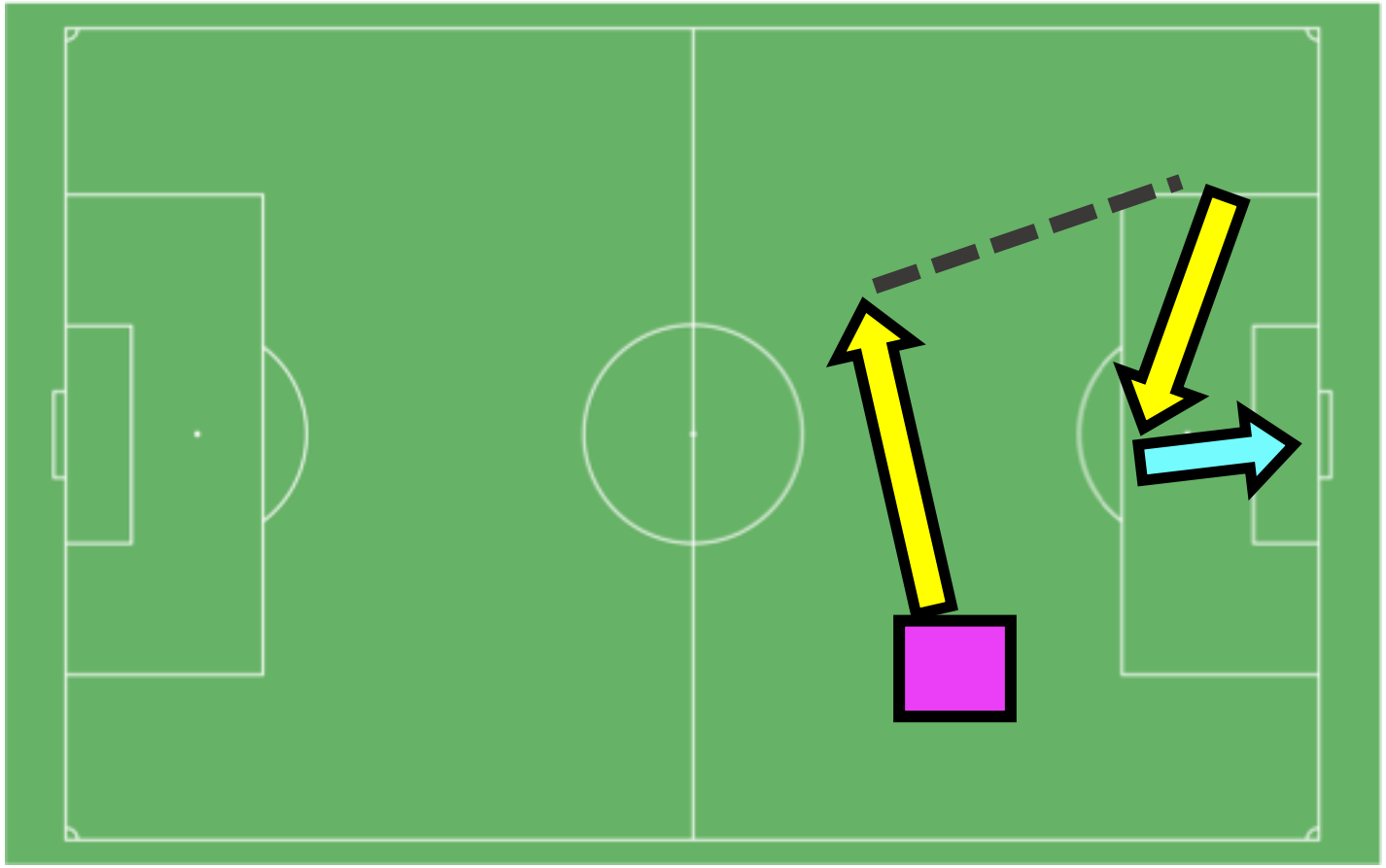
Tactics involve space and time

~~Frequent itemsets : {Pass, Goal, Pass, Dribble, Tackle}~~

Tactics involve space and time

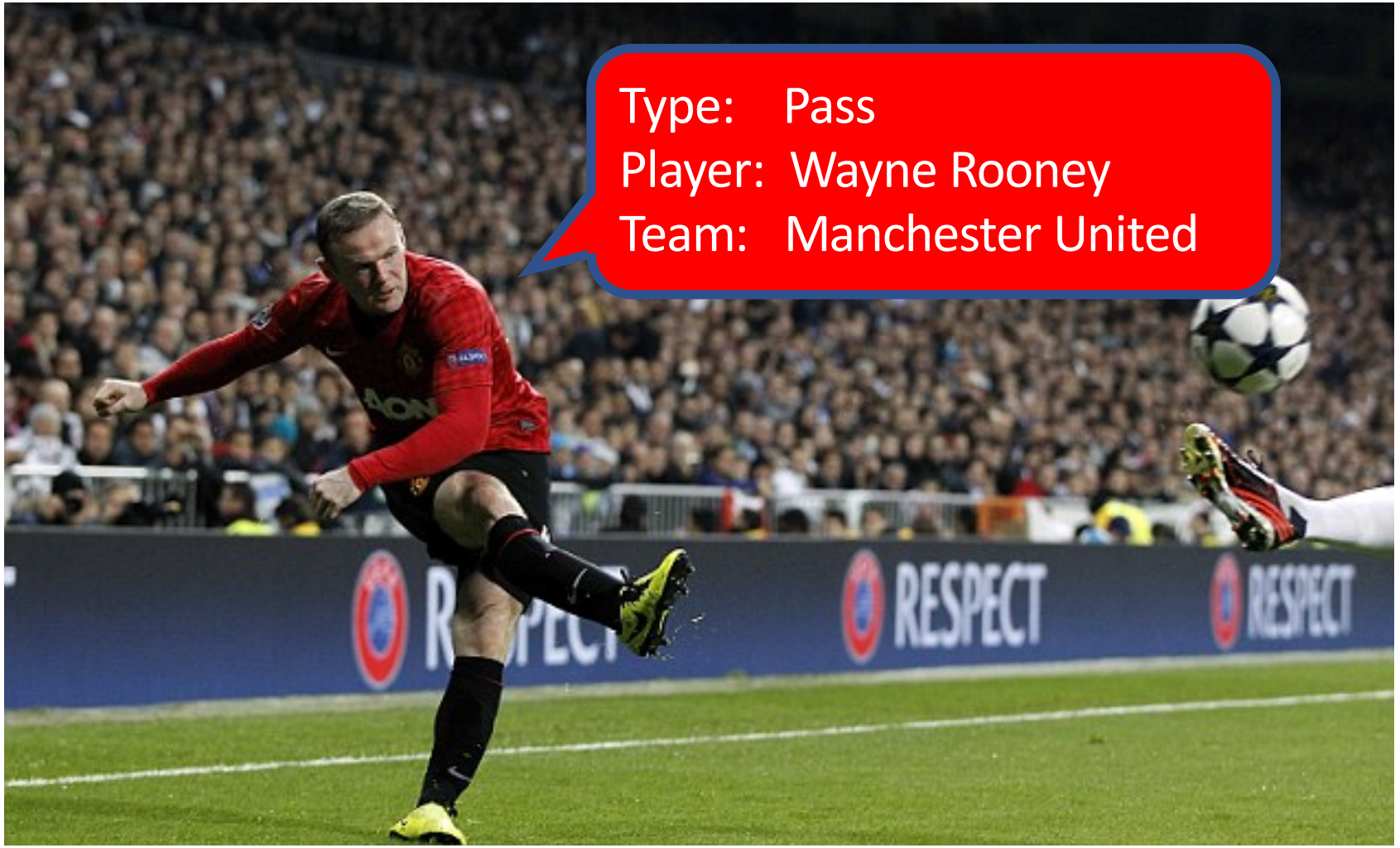
~~Frequent itemsets : {Pass, Goal, Pass, Dribble, Tackle}~~

Tackle
⇒ Pass
⇒ Dribble
⇒ Pass
⇒ Goal

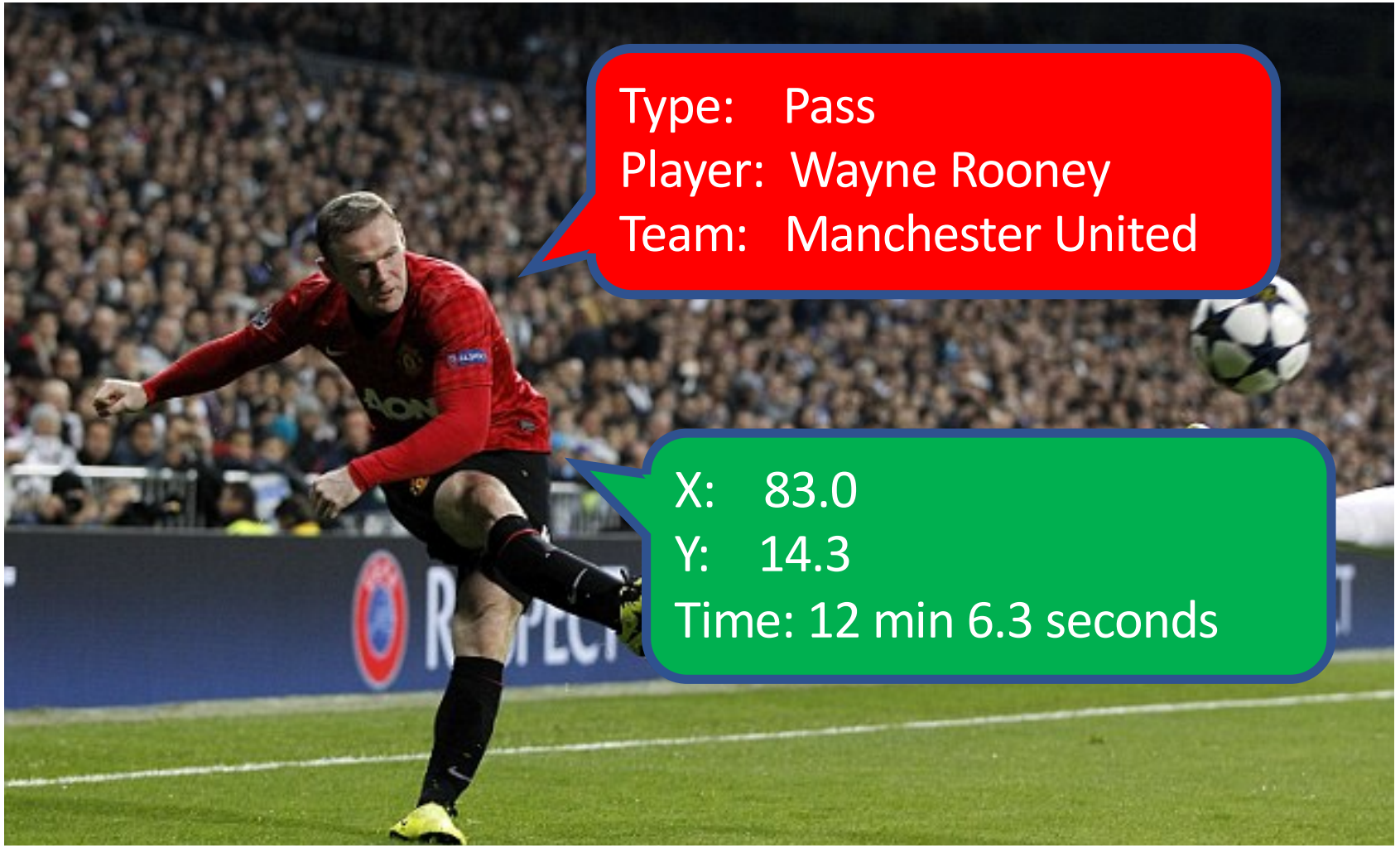


Events have discrete and continuous attributes

Events have discrete and continuous attributes



Events have discrete and continuous attributes



How to exploit domain knowledge?

Raw data

- Player: Hazard
- Type: Pass

How to exploit domain knowledge?

Enriched raw data

- Player: Hazard -> Left Winger -> Attacker
- Type: Pass -> Corner, Cross

Relevance of tactics is subjective

Attackers care about defence patterns

Defenders care about attack patterns

Coaches care about successful patterns

Journalists care about unique patterns

Events have no universal definition

Events have no universal definition

Aggressive cross?

Events have no universal definition

Aggressive cross?

OR off-target shot?

Automatic discovery of tactics in spatio-temporal soccer match data

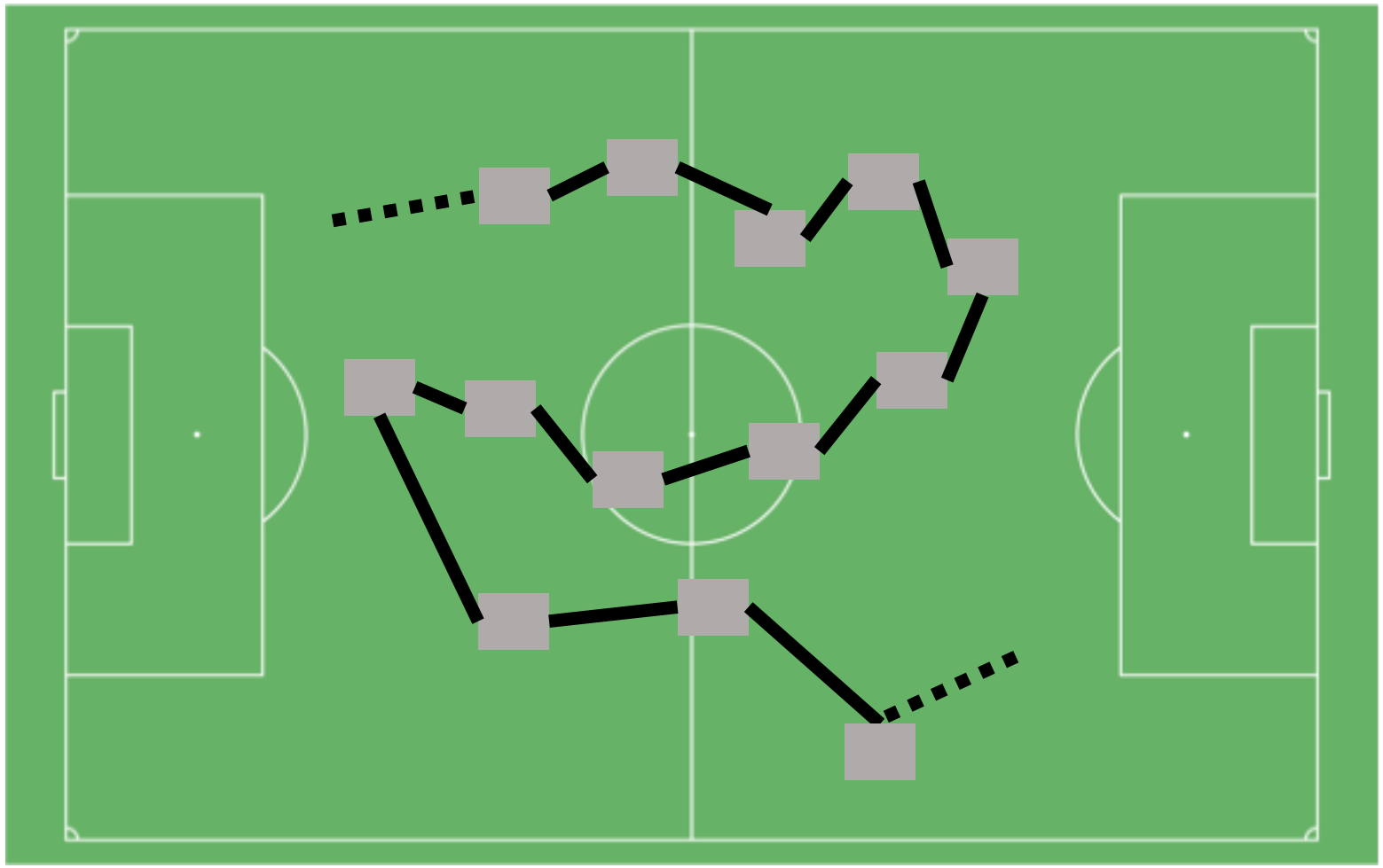
Data

Challenges

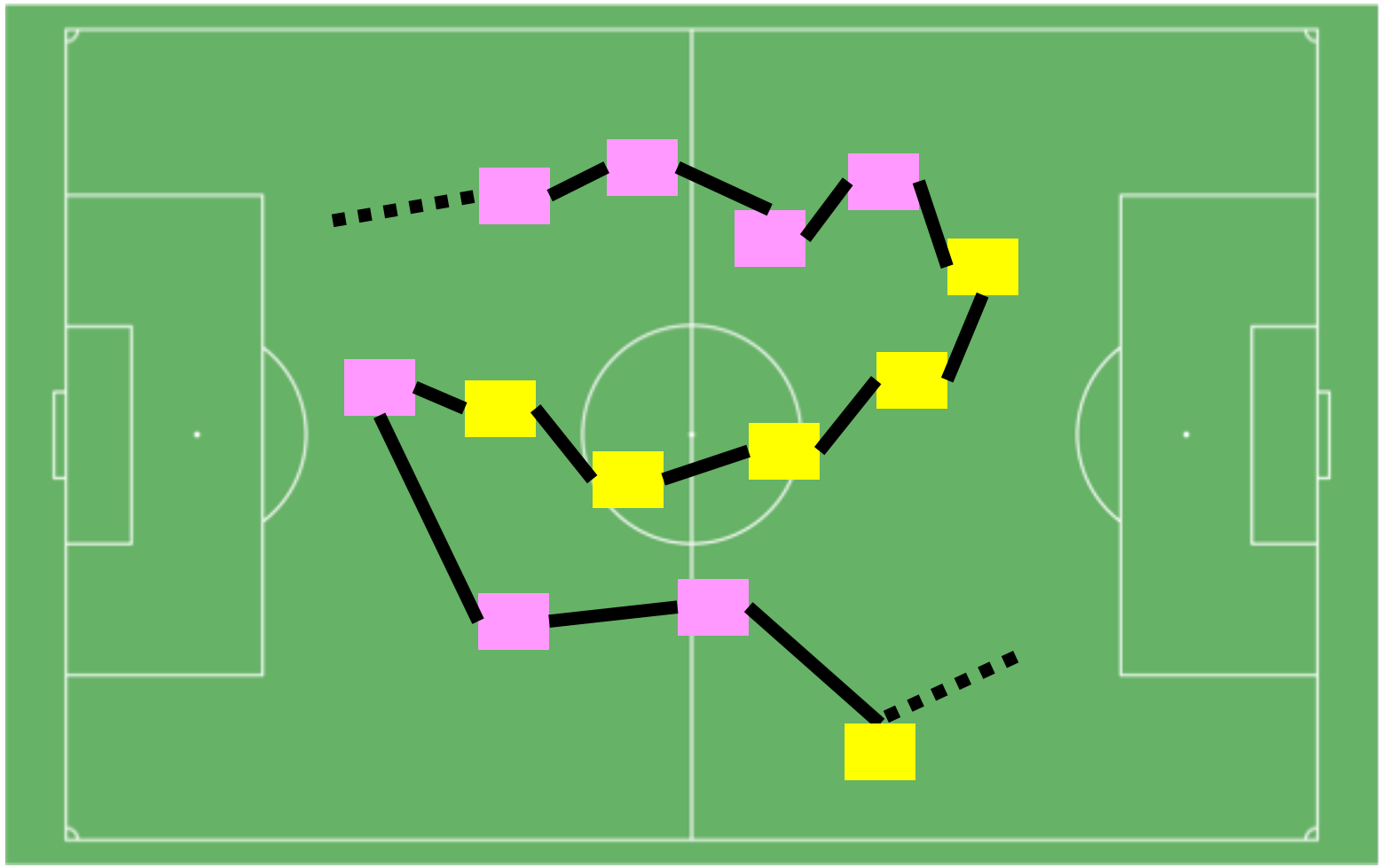
Approach

Results

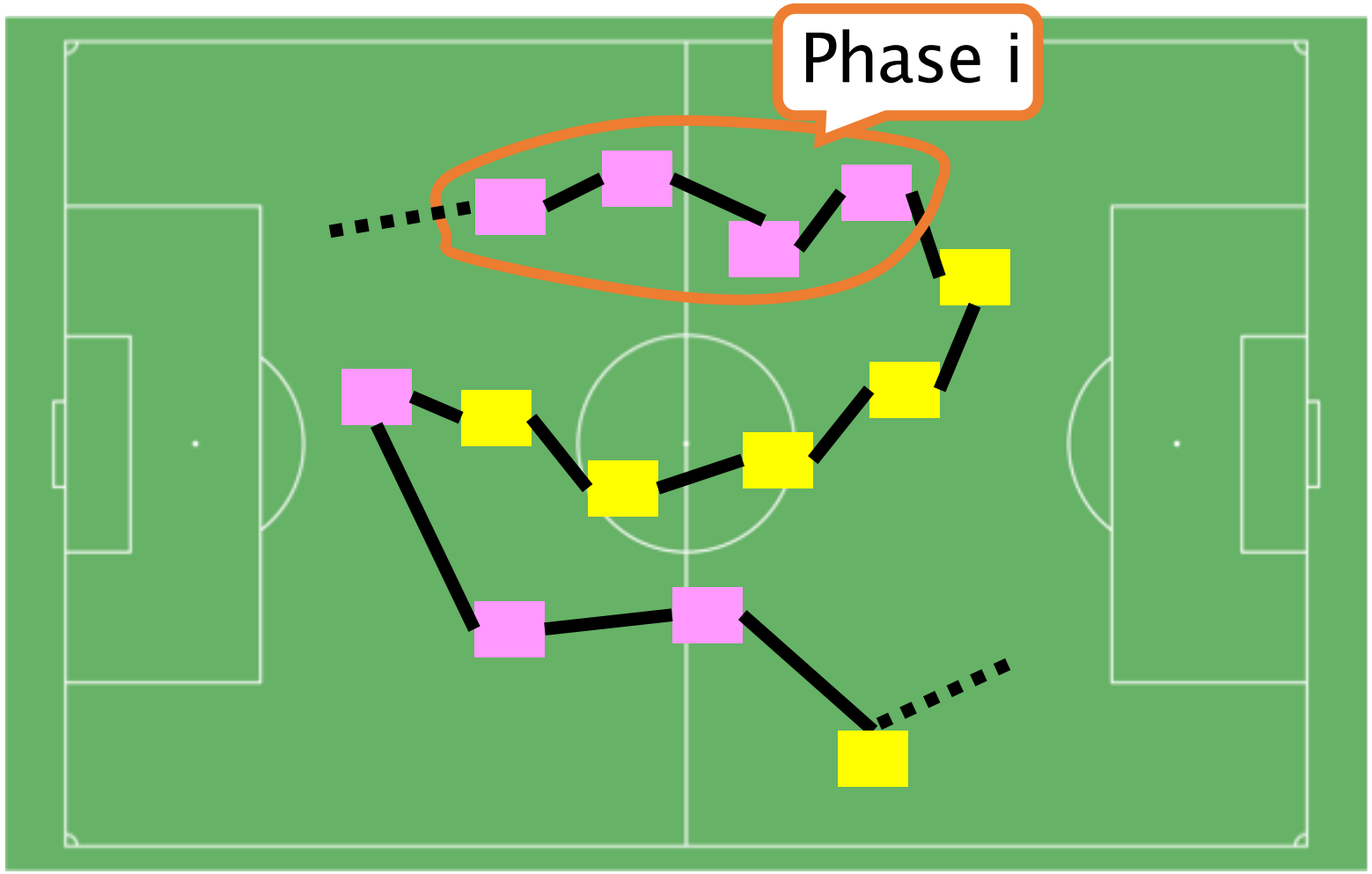
Step 1: Divide the event stream of each match into phases



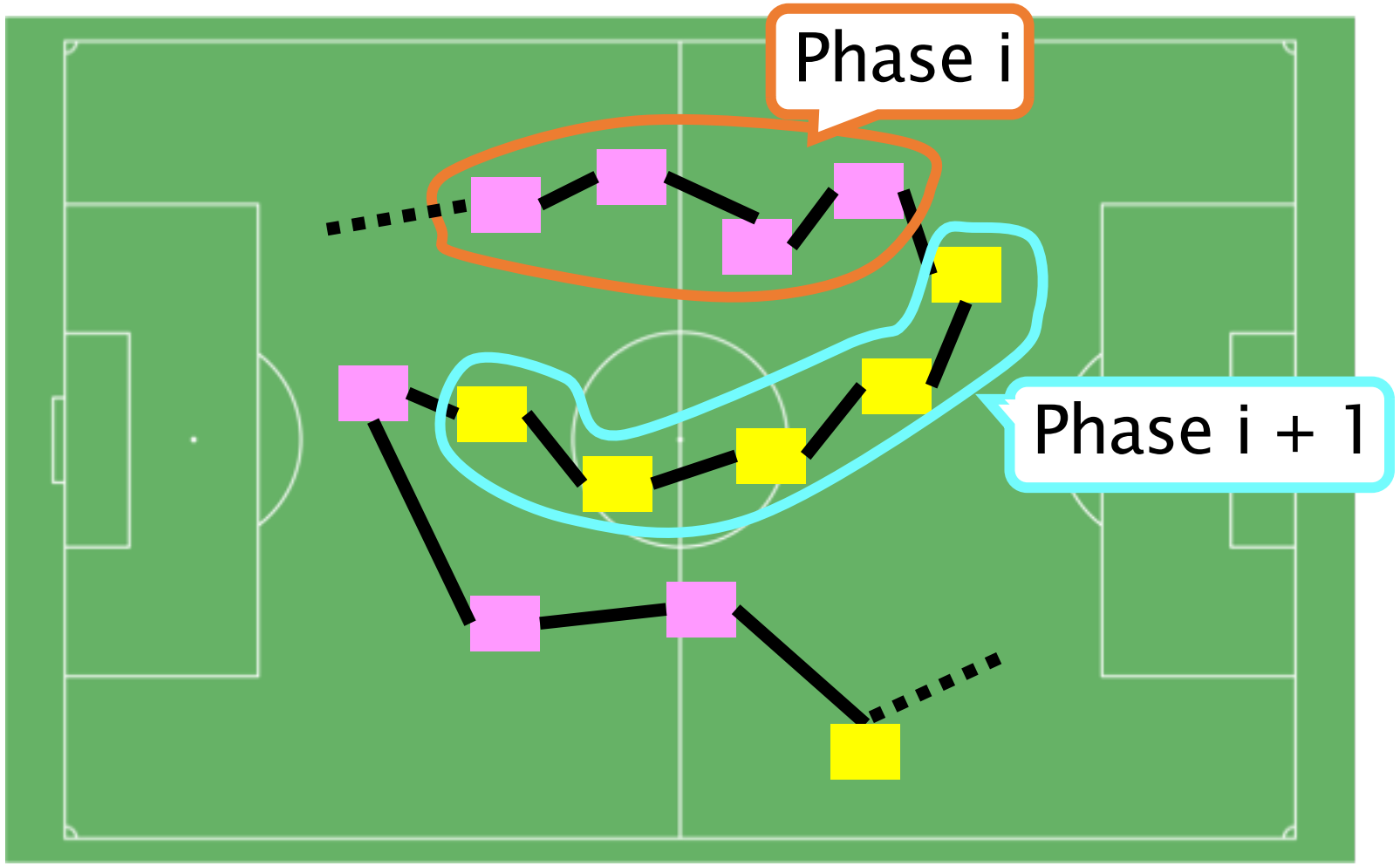
Step 1: Divide the event stream of each match into phases



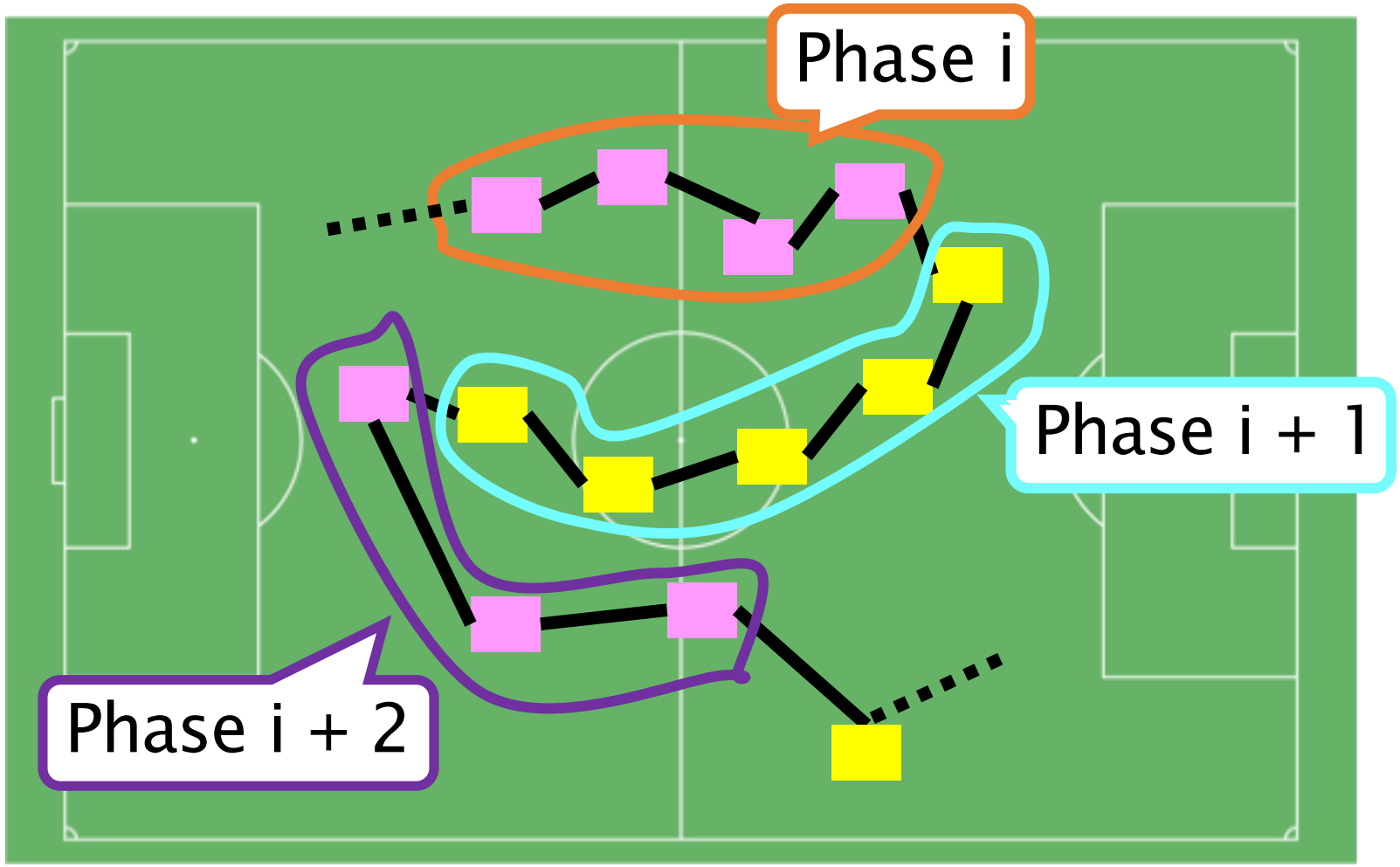
Step 1: Divide the event stream of each match into phases



Step 1: Divide the event stream of each match into phases



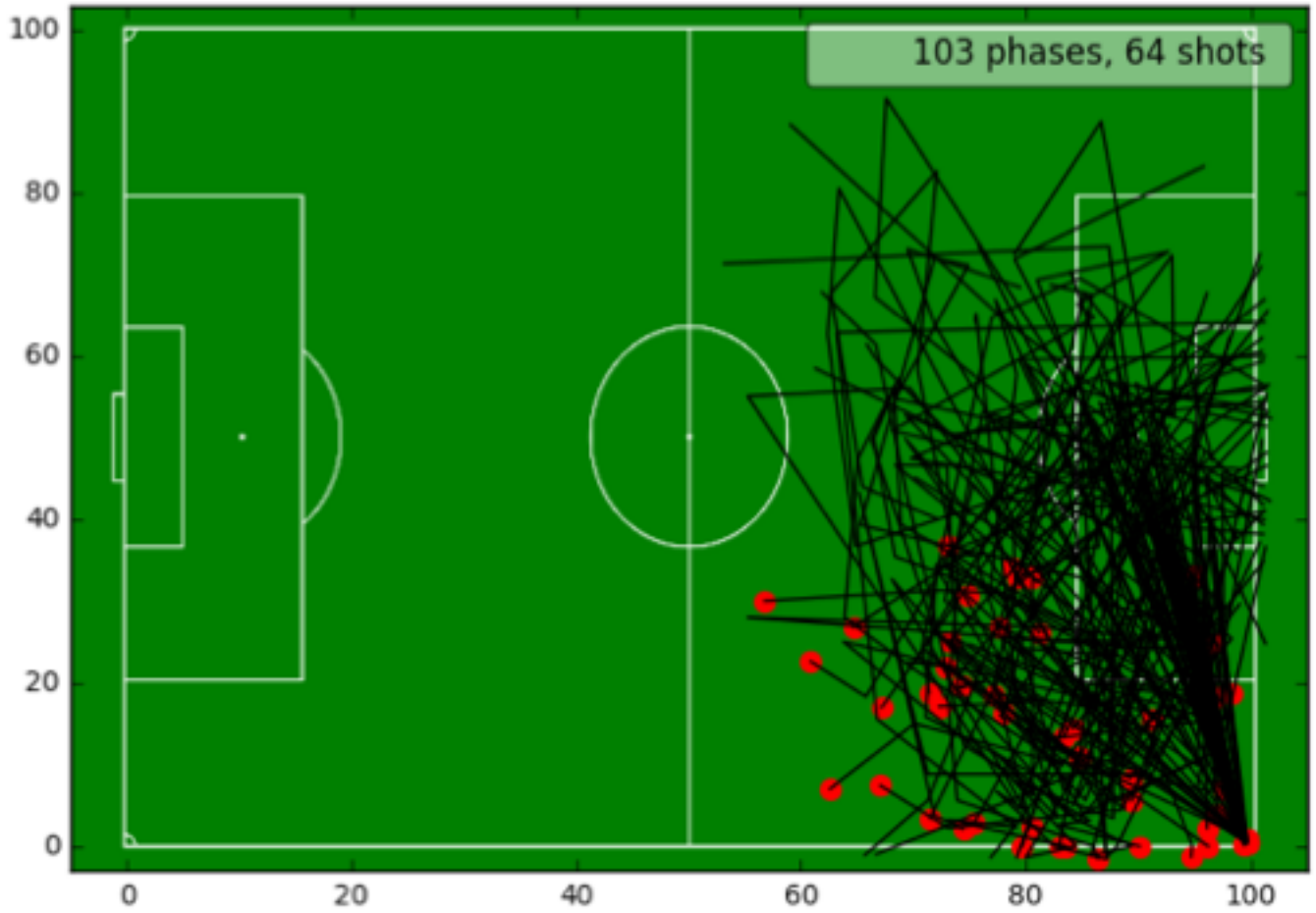
Step 1: Divide the event stream of each match into phases



Step 2: Cluster phases on their spatio-temporal component

Distance function =
Dynamic Time Warping

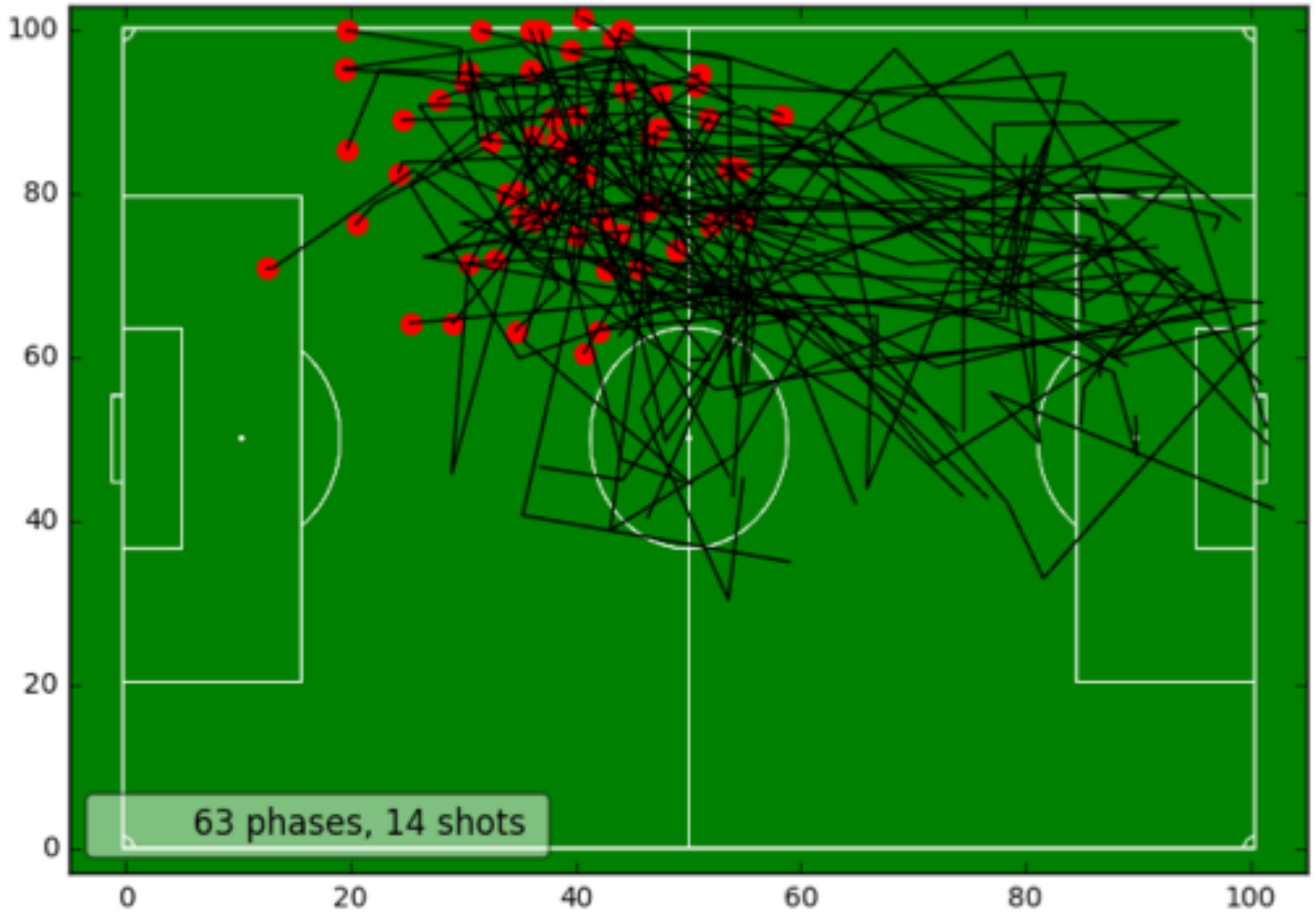
Parameter =
How many clusters?



Step 2: Cluster phases on their spatio-temporal component

Distance function =
Dynamic Time Warping

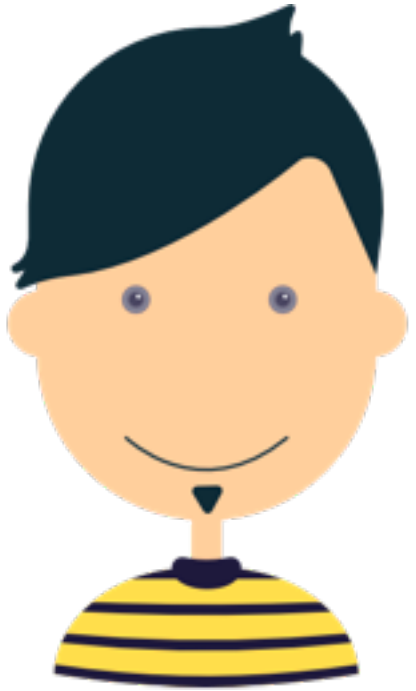
Parameter =
How many clusters?



Step 3: Rank clusters on user preference

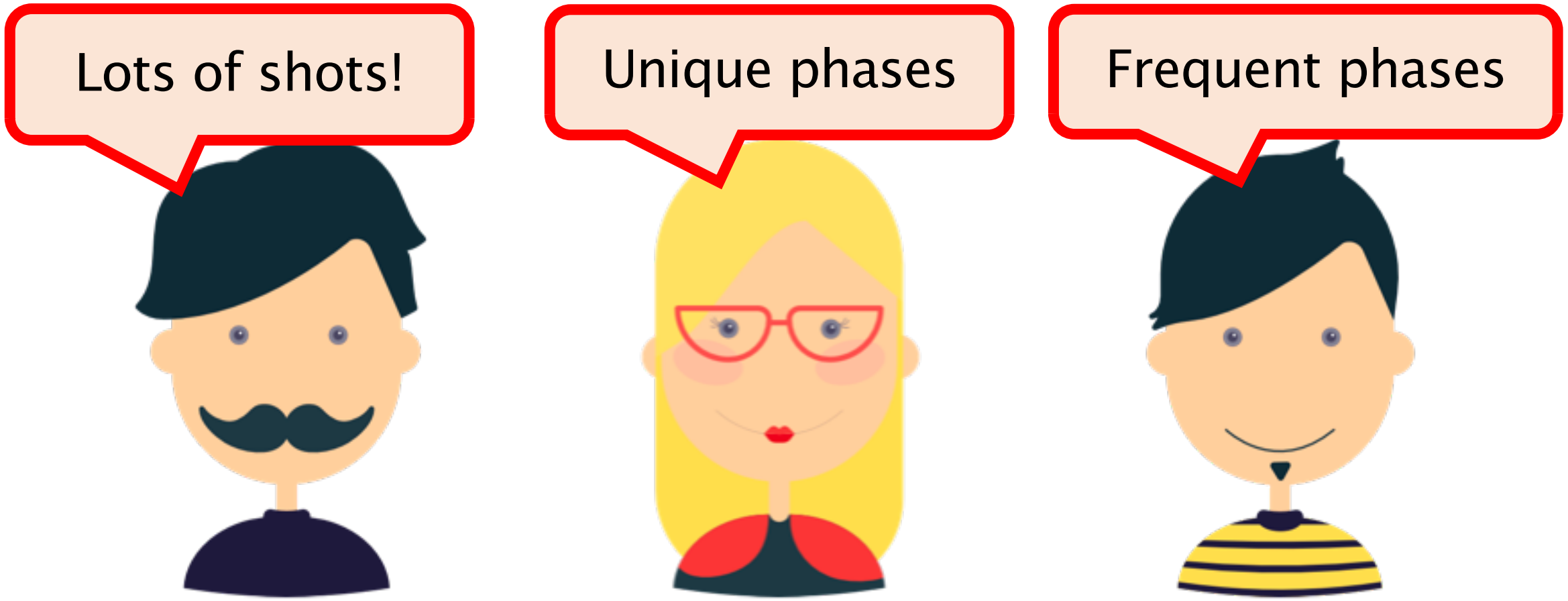
Soccer coach

Journalist



Opposing player

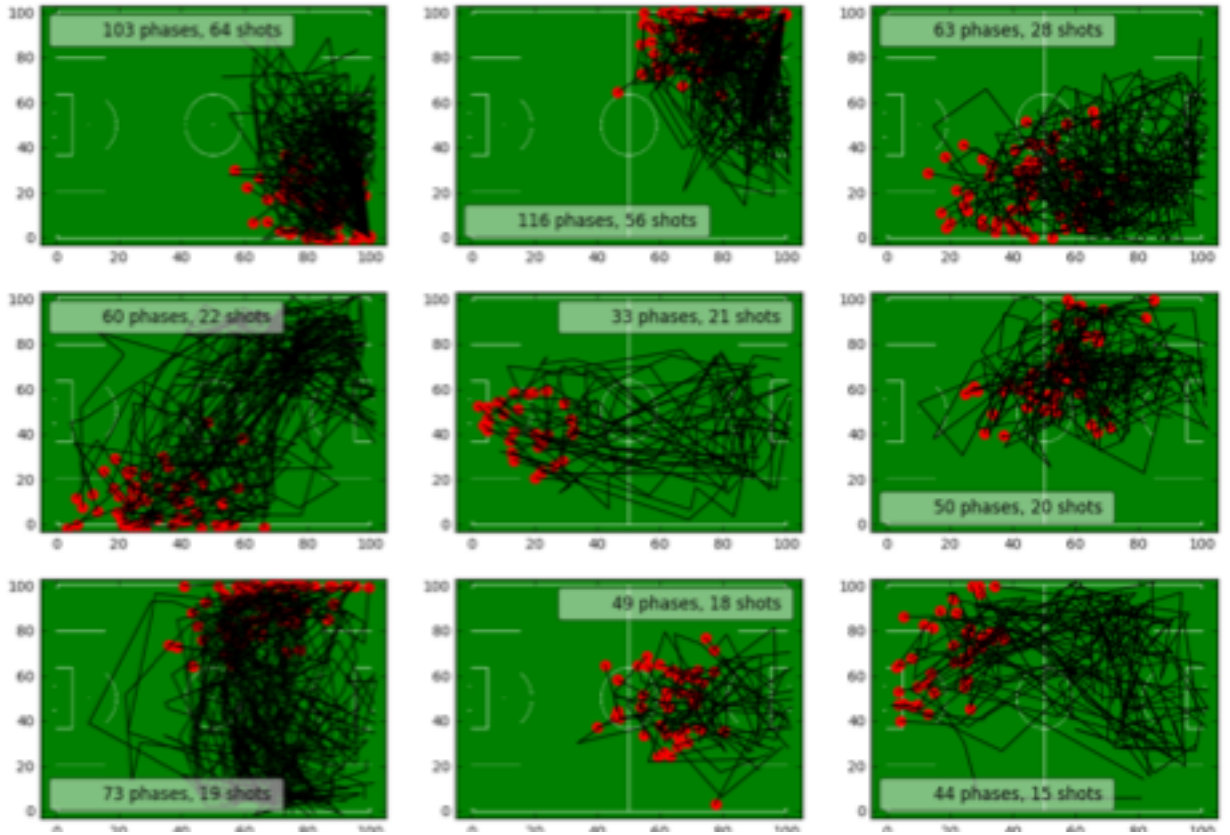
Step 3: Rank clusters on user preference



Step 3: Rank clusters on user preference

Soccer coach

9 clusters with most shots



Step 4: Mine each cluster for frequent sequential patterns

Sequential pattern mining algorithms (e.g. CM-SPADE) work on sequences of itemsets

Step 4: Mine each cluster for frequent sequential patterns

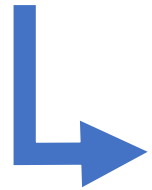
Sequential pattern mining algorithms (e.g. CM-SPADE) work on sequences of itemsets

Q: How to convert a soccer event to an itemset?

Step 4: Mine each cluster for frequent sequential patterns

Sequential pattern mining algorithms (e.g. CM-SPADE) work on sequences of itemsets

Q: How to convert a soccer event to an itemset?



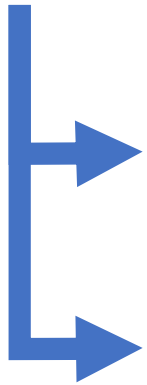
q1: What information to consider?

Player/Event type/Location

Step 4: Mine each cluster for frequent sequential patterns

Sequential pattern mining algorithms (e.g. CM-SPADE) work on sequences of itemsets

Q: How to convert a soccer event to an itemset?



q1: What information to consider?

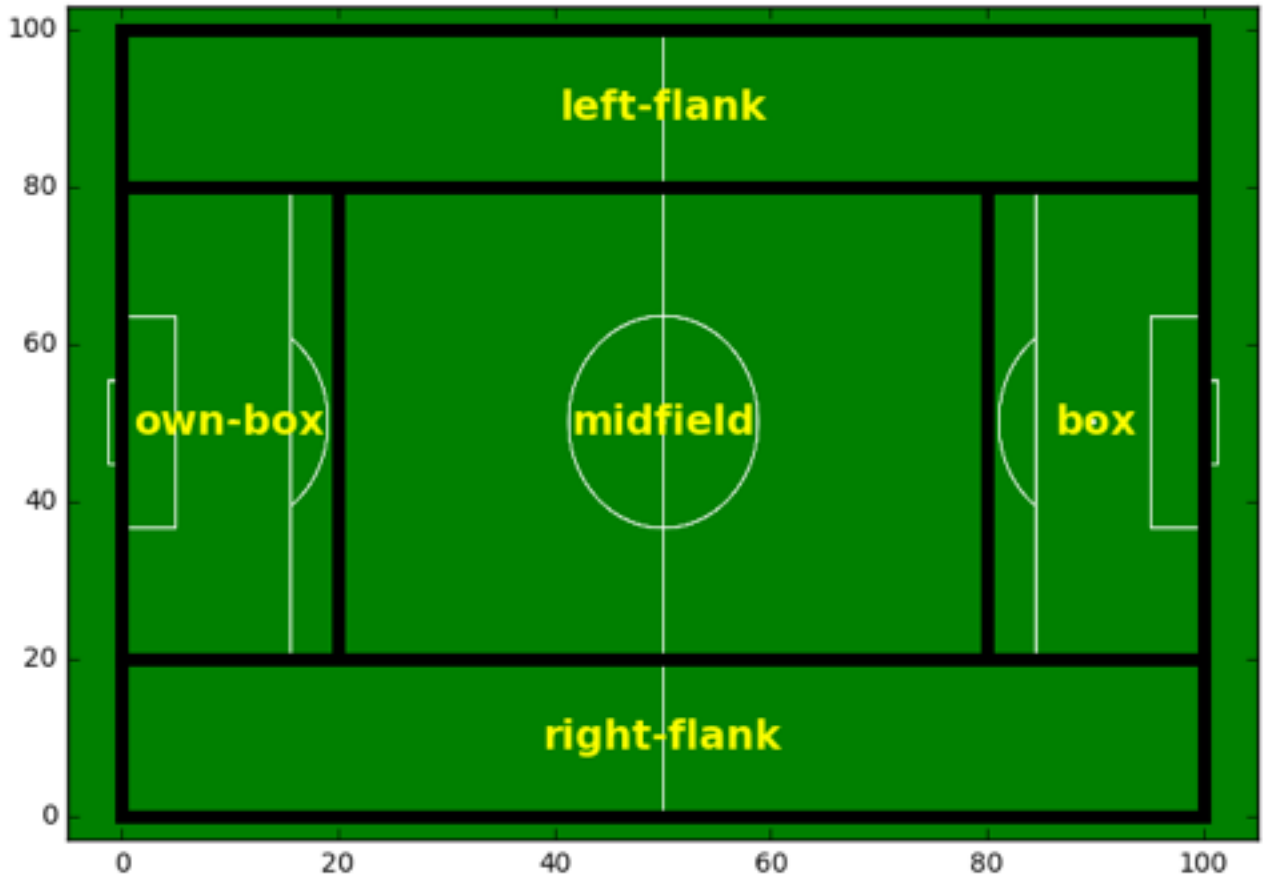
Player/Event type/Location

q2: How to encode this information?

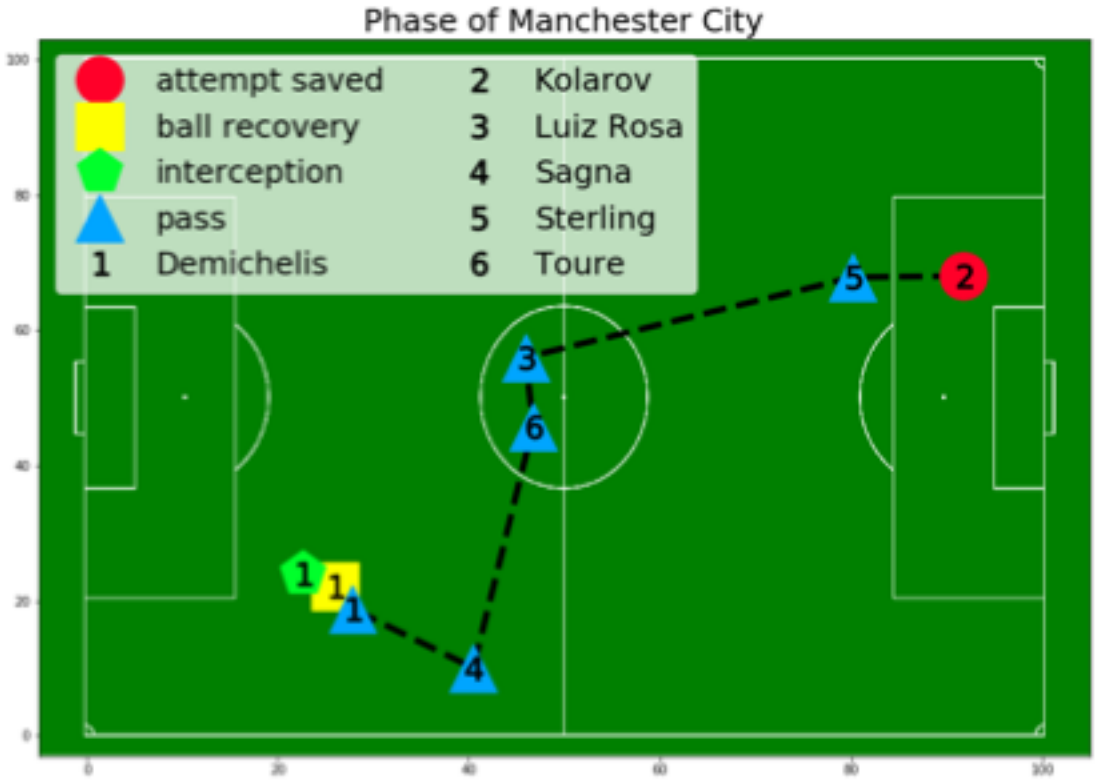
Discretization? Hierarchy?

Step 4: Mine each cluster for frequent sequential patterns

Location discretization
based on domain knowledge

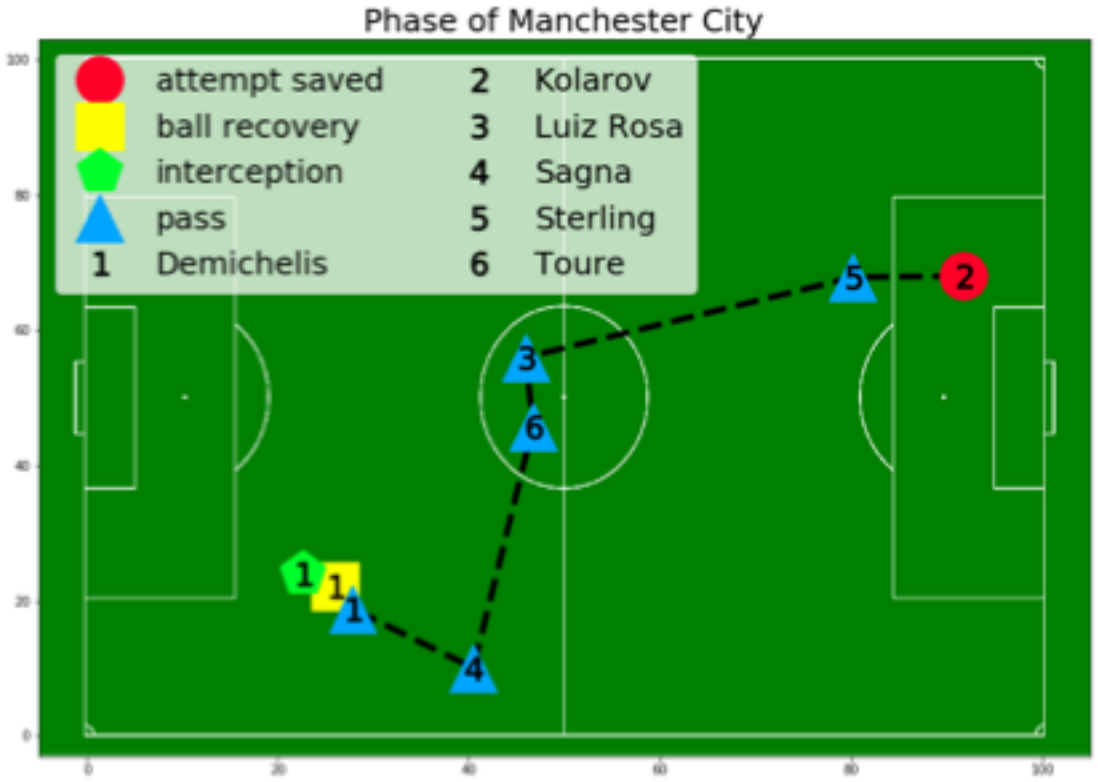


Step 4: Mine each cluster for frequent sequential patterns



1. Interception AT the right flank
2. Ball recovery AT the right flank
3. A pass FROM the right flank TO the right flank
4. A pass FROM the right flank TO the midfield
5. A pass FROM the midfield TO the midfield
6. A pass OR long ball FROM the midfield TO the box
7. A pass FROM the box TO the box
8. Attempt saved AT the box
9. Shot AT box

Step 4: Mine each cluster for frequent sequential patterns



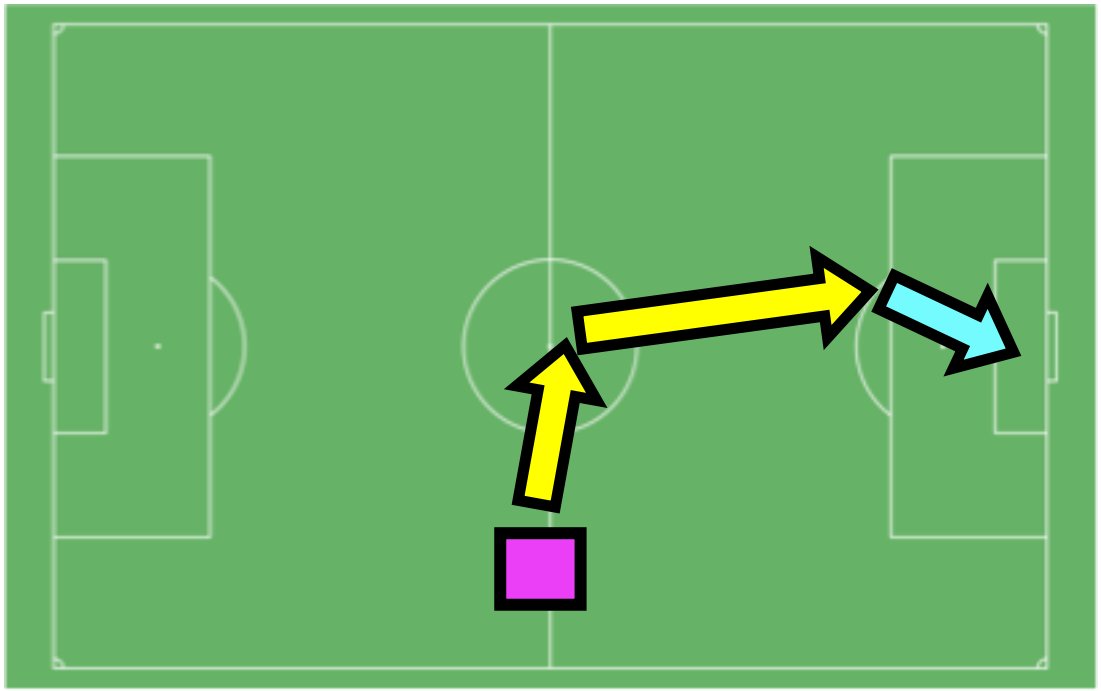
2. Ball recovery AT the right flank

4. A pass FROM the right flank TO the midfield

6. A pass OR long ball FROM the midfield TO the box

9. Shot AT the box

Step 4: Mine each cluster for frequent sequential patterns



Ball recovery AT the right flank

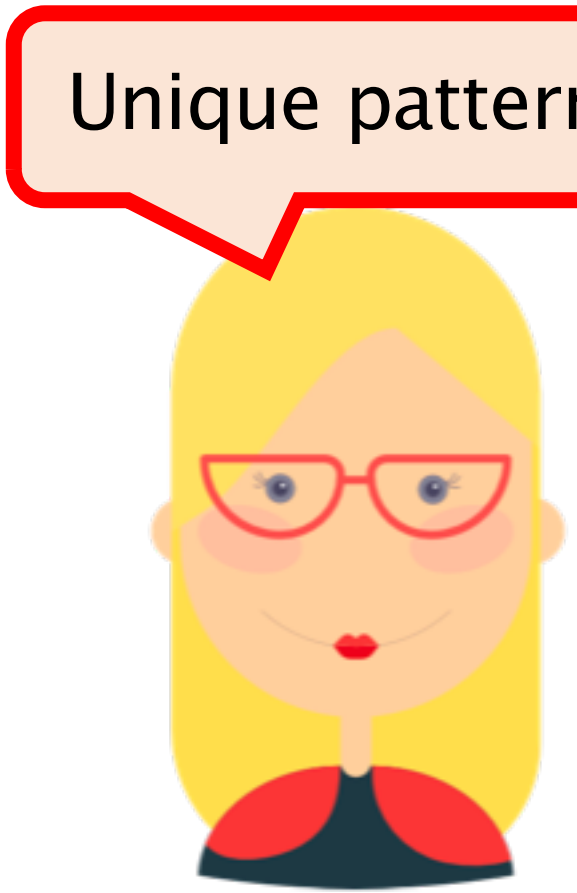
⇒ A pass FROM the right flank TO the midfield

⇒ A long ball FROM the midfield TO the box

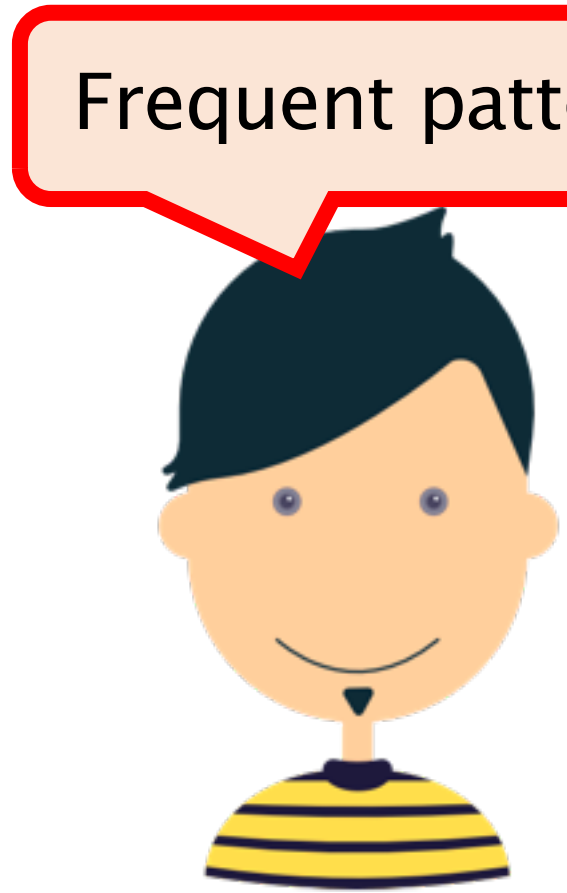
⇒ Shot AT the box

Step 5: Rank patterns on user preference

Soccer coach



Journalist



Opposing player

Step 5: Rank patterns on user preference

Soccer coach

Step 5: Rank patterns on user preference

Soccer coach

$W_{ball_recovery} = 1$

$W_{pass} = 0.5$

$W_{shot} = 2$

Step 5: Rank patterns on user preference

Soccer coach

Ball recovery AT the right flank

⇒ A pass FROM the right flank TO the midfield

⇒ A long ball FROM the midfield TO the box

⇒ Shot AT the box

$W_{ball_recovery} = 1$

$W_{pass} = 0.5$

$W_{shot} = 2$

Step 5: Rank patterns on user preference

Soccer coach

Ball recovery AT the right flank
⇒ A pass FROM the right flank TO the midfield
⇒ A long ball FROM the midfield TO the box
⇒ Shot AT the box

$W_{ball_recovery} = 1$
$W_{pass} = 0.5$
$W_{shot} = 2$

$$Score(p) = Support(p) * (1 + 0.5 + 0.5 + 2)$$

Automatic discovery of tactics in spatio-temporal soccer match data

Data

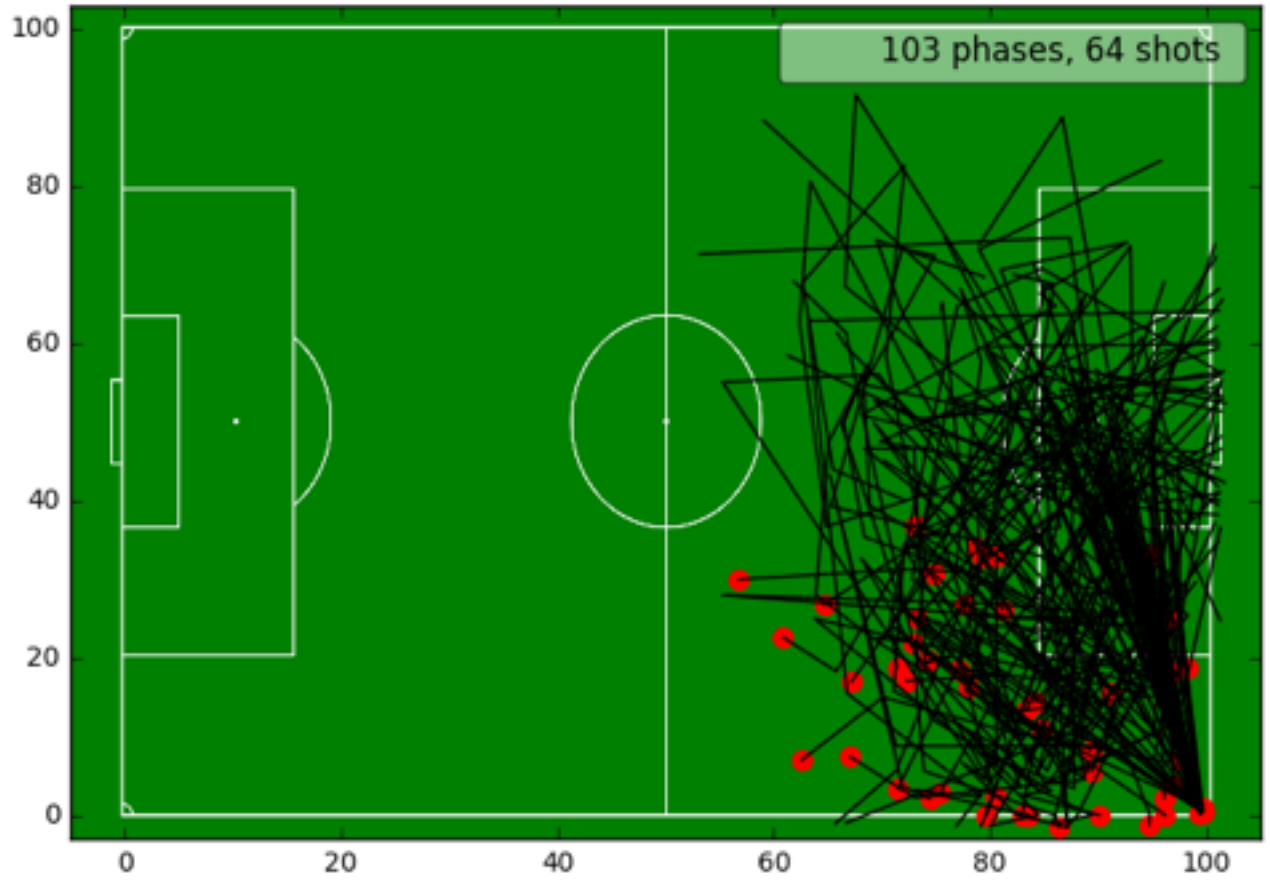
Challenges

Approach

Results

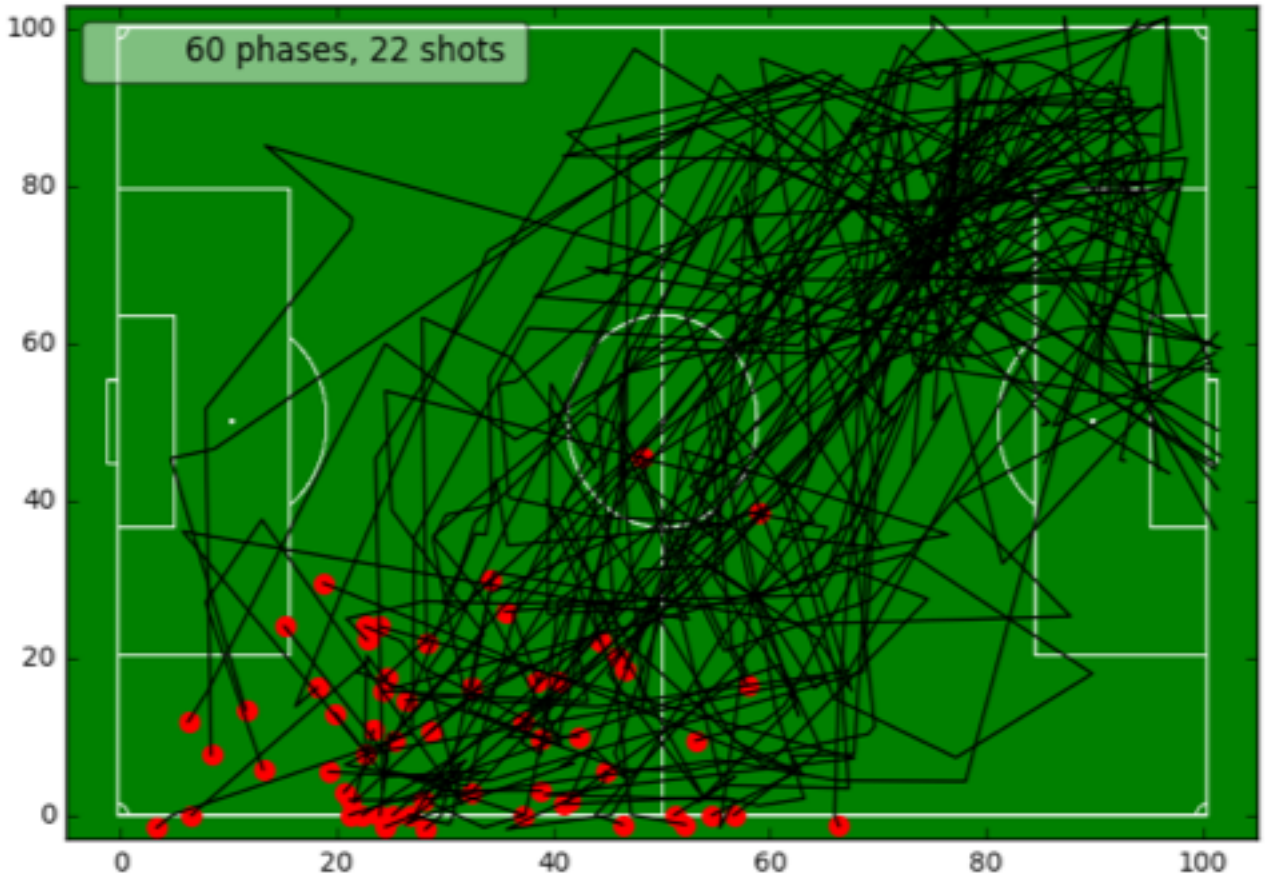
Q1: Do we discover interesting and relevant patterns?

Top ranked cluster of Manchester City in 2015/2016



Q1: Do we discover interesting and relevant patterns?

4th ranked cluster of Manchester City in 2015/2016



Q2: Can we identify team tactics?

Top ranked patterns in clusters of

Arsenal

Cluster	Sequential Pattern
2^{nd} Cluster	1. A pass OR cross FROM the left flank TO the box 2. Shot
3^{rd} Cluster	1. A pass FROM the midfield TO the midfield 2. A pass FROM the midfield TO the midfield 3. A pass FROM the midfield TO the midfield
9^{th} Cluster	1. A pass FROM the midfield TO the midfield 2. A pass FROM the midfield TO the left-flank 3. A pass FROM the left flank TO the midfield 4. A pass FROM the midfield TO the midfield 5. A pass FROM the midfield TO the midfield

Q2: Can we identify team tactics?

Top ranked patterns in clusters of

Leicester City

Cluster	Sequential Pattern
1^{st} Cluster	1. A pass OR cross FROM the left flank TO the box 2. A shot
2^{nd} Cluster	1. A pass OR cross FROM the right flank TO the box 2. A shot and a Miss 3. Ball goes out of bounds
7^{th} Cluster	1. A ball recovery IN the midfield 2. A shot

Conclusion

Have we made Bob’s job easier?

Conclusion

Have we made Bob's job easier?

Not yet, but it's a good start!