Qualitative spatial reasoning for soccer pass prediction

Vincent Vercruyssen
University of Leuven, Belgium

September 19, 2016

Contents

Contents

Qualitative spatial reasoning

Suppose we have spatiotemporal data.

Hypothesis:

It is possible to learn a meaningful qualititative model over the data

How to test this?

Soccer pass prediction based on spatiotemporal player data:

“Can we predict to whom a player is going to give a pass?”

Contents

Soccer match data

- During a soccer match, three, different types of data are available

1. Spatiotemporal data

player_ID	time	X	Y	events_half
345555	18500	-3455	300	1
356778	18500	220	-1567	1
245777	18500	10	-908	2

2. Event data

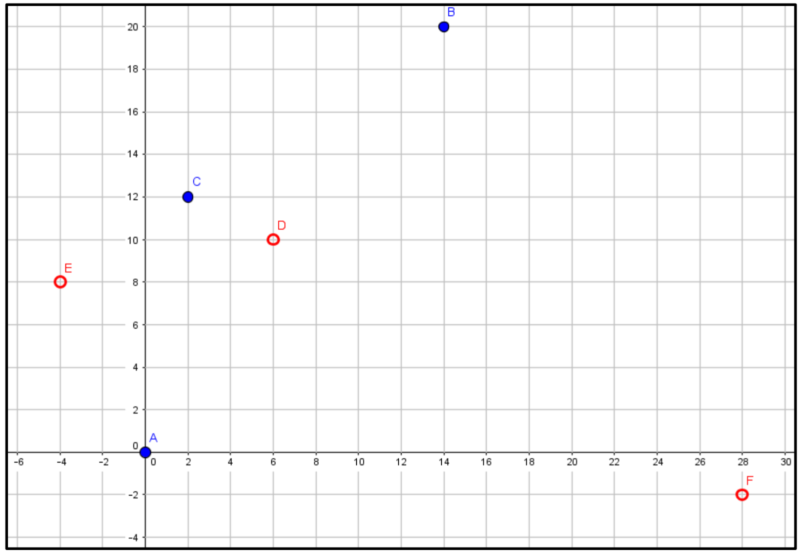
player_ID	time	event	…	events_half
345555	18500	pass	…	1
356778	18500	reception	…	1
245777	22300	pass	…	2

3. Background knowledge

player_ID	team	position	…	name
345555	A	midfield	…	Jack
356778	A	defender	…	Stephen
245777	B	attack	…	John

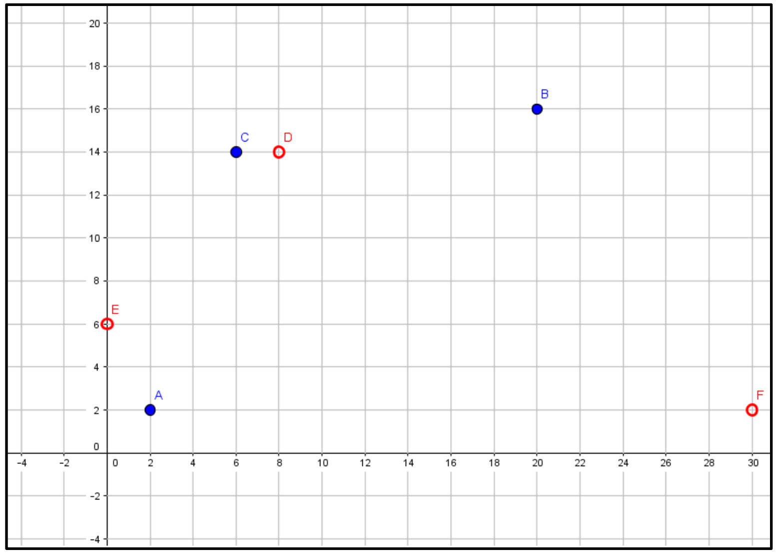
Pass event

t-2 (no pass)



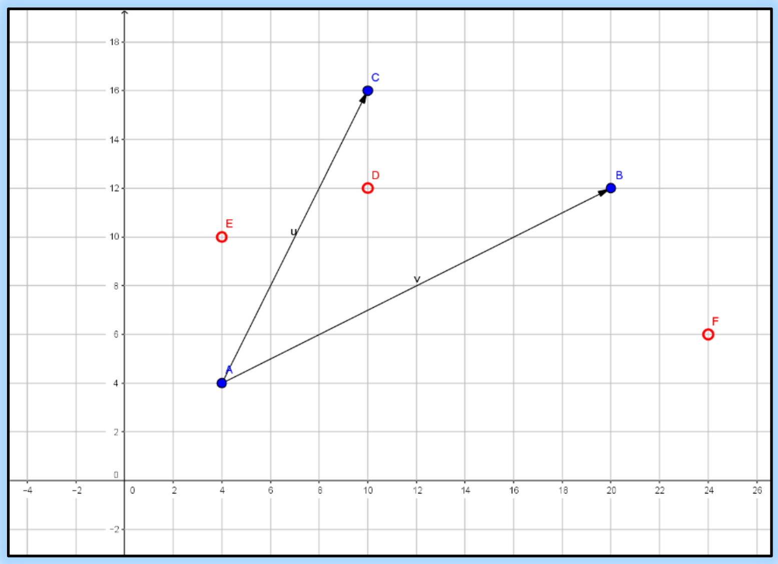
	A	B	C	D	E	F
X	0	14	2	6	-4	28
Y	0	20	12	10	8	-2

t-1 (no pass)



	A	B	C	D	E	F
X	2	20	6	8	0	30
Y	2	16	14	14	6	2

t (pass)



	A	B	C	D	E	F
X	4	20	10	10	4	24
Y	4	12	16	12	10	6

Contents

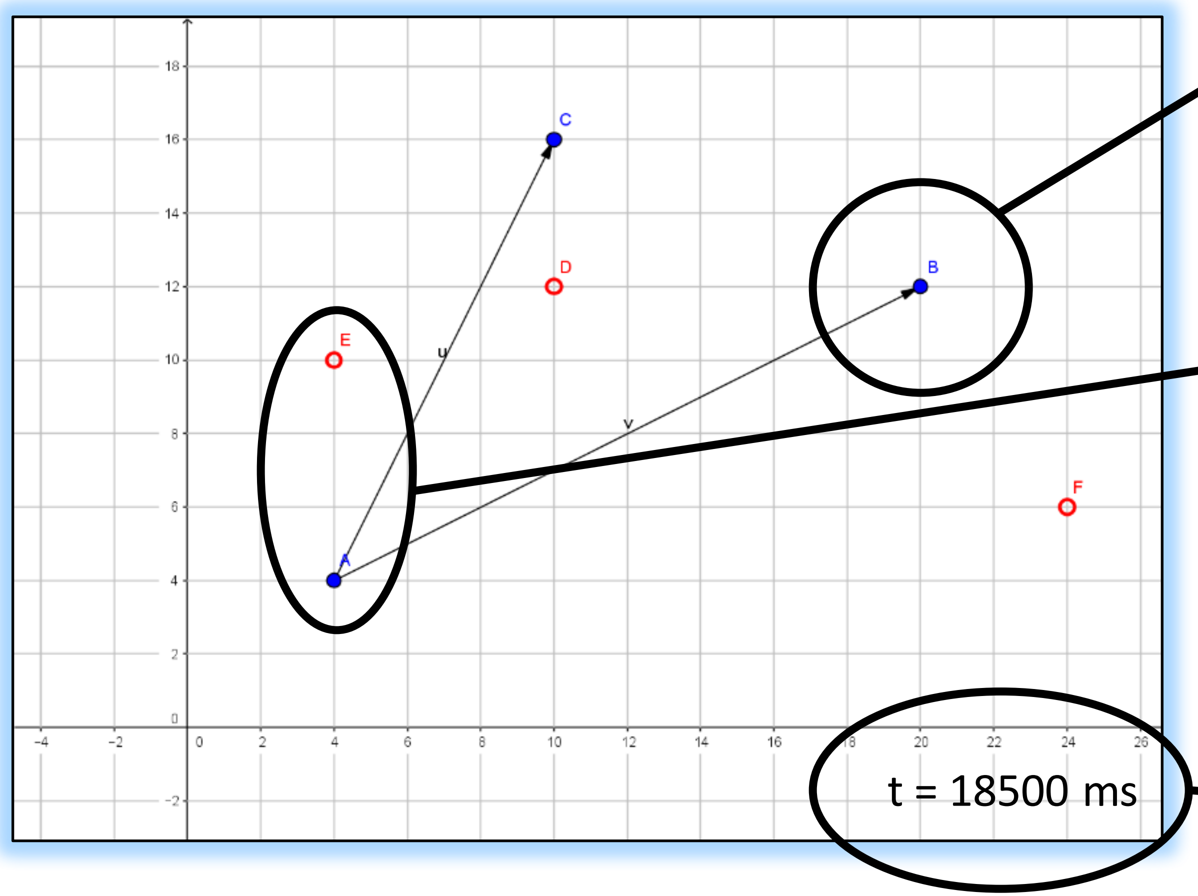
Quantitative reasoning...

- Difficult to learn directly over exact spatiotemporal data
  - No single pass will be given in the same exact locations
  - Size of the pitch will change between stadiums = different reference framework
  - Prone to inaccurate measurements

- Soccer data contain relations and complex interactions
  - players base their decisions on how they are positioned with respect to other players...
  - ...and how these players interact

- Soccer data are inherently dynamic
  - passing decisions are made in the moments leading up to the pass

Challenges: pass event



The exact position will never be the same

How can we express relations between players?

$$player(A, E, north)$$

$$player(B, free)$$

What about the moments leading up to the pass?

… or qualitative reasoning?

- Difficult to learn directly over exact spatiotemporal data

  → generalization

- Soccer data contain relations and complex interactions

  → framework to express relations + combine different types of knowledge

- Soccer data are inherently dynamic

  → encode information over time

Contents

Methodology

Goal: learn a predictive model from data

1. **Data**: consider each pass event as a labelled training example
   - Positive example = player that receives the pass
   - Negative example = other teammembers on the field at that time
2. **Features**: extract features that qualitatively describe the pass event
3. **Model**: Learn a prediction model using features and background info
4. **Predict**: Construct ranking of who is most likely to receive a pass in unseen example

Extract qualitative features

- Qualitative spatial reasoning (QSR) is an umbrella term for a number of formalisms (calculi) that define how entities in a 2D or 3D space behave

- QSR's describe relations between objects in a qualitative way

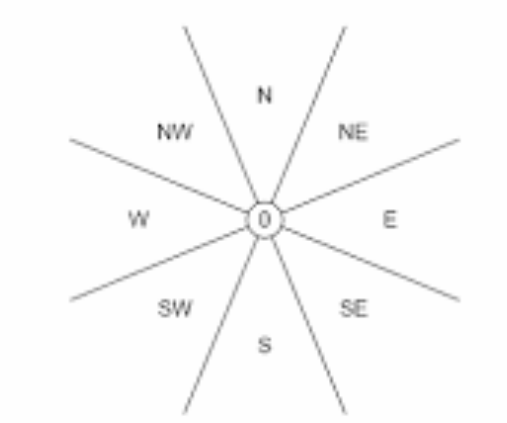
- Relations are mostly binary, yet can have higher degrees

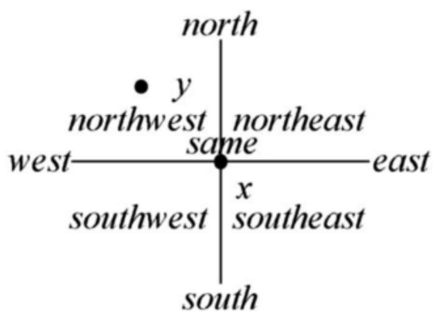
- Numerous categories of QSR's exist:
  - Mereotopology
  - Direction
  - Distance
  - Moving objects
  - Shape
  - ...

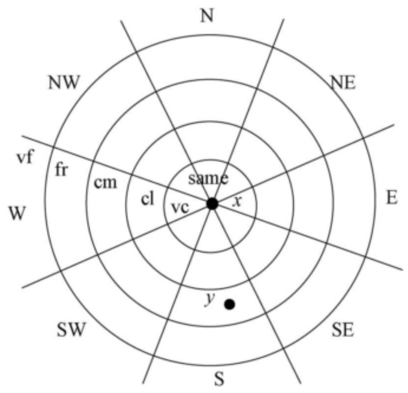
These are interesting for the problem at hand

Qualitative Spatial Representations

- **Cone-shaped direction calculus** OR **projection-based direction calculus**
  - 8 binary relations – JEPD (*jointly exhaustive pairwise disjoint*)
  - These basic calculi can be extended with distance information
  - Represents static relations





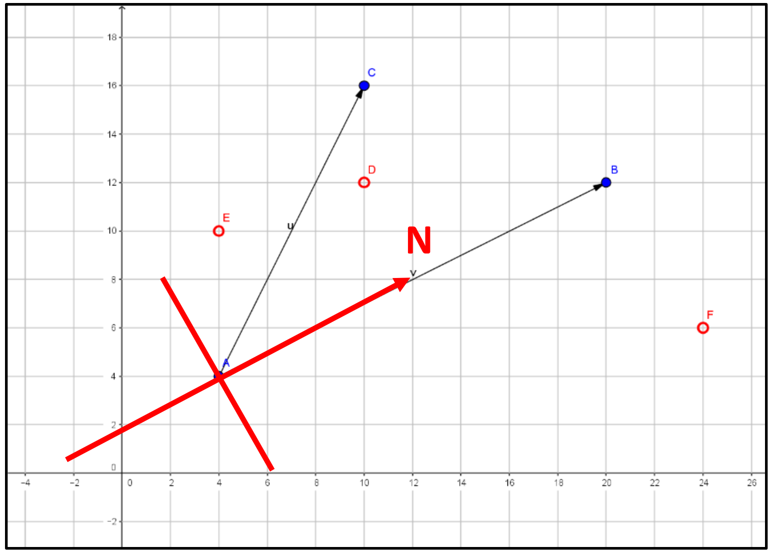


Directional information

Directional and distance information

Qualitative Spatial Representations

- **Cone-shaped direction calculus** OR **projection-based direction calculus**
  - Use the *receiver* and *passer* as points of reference
  - Capture players' position with regards to passer and receiver



	A	B	C	D	E	F
A		**N**	**NW**	**NW**	**W**	**NE**
B	**S**		W	SW	SW	E
C	**SE**	E		E	S	E
D	**SE**	NE	W		S	E
E	**E**	NE	N	N		NE
F	**SW**	W	W	W	SW	

passer

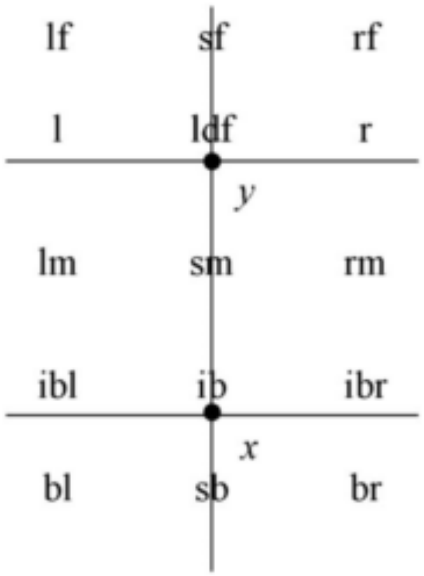
actual receiver

No pass (from A to C):

	NE	**N**	**N**	**N**	**NE**

Qualitative Spatial Representations

- **Double-cross calculus** OR **LR calculus**
  - 15 *ternary* JEPD relations
  - Represents static relations

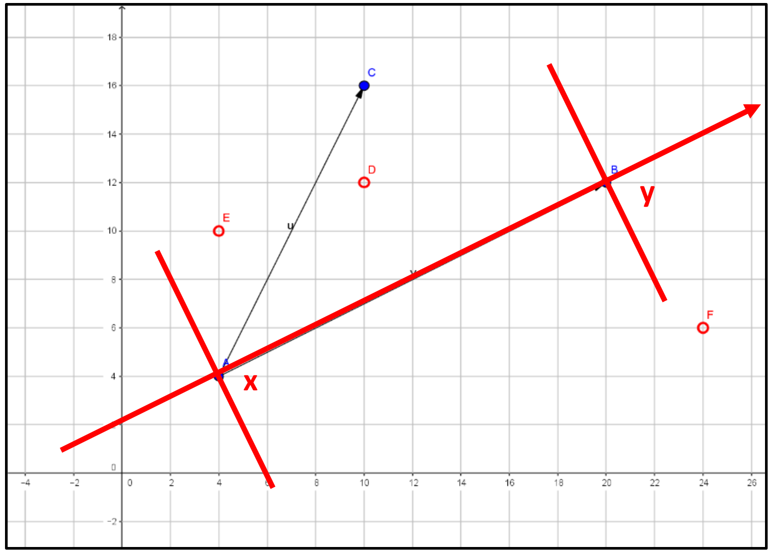


Double-cross

LR calculus

Qualitative Spatial Representations

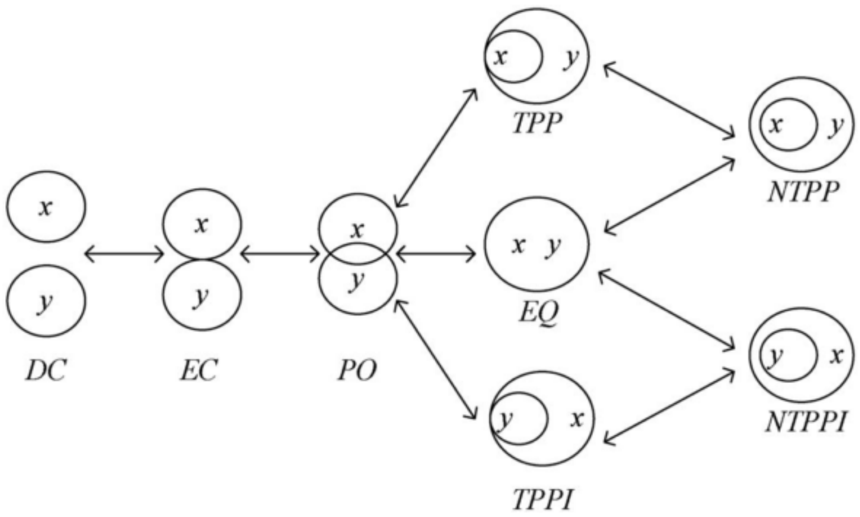
- **Double-cross calculus** OR **LR calculus**
  - Use the *passline* as a point of reference
  - Captures players’ position with regards to the passline



	A	B	C	D	E	F
ref	ib	rf	ldf	rm	lm	rm
ref	ib	ldf	lm	lm	lm	rf

Qualitative Spatial Representations

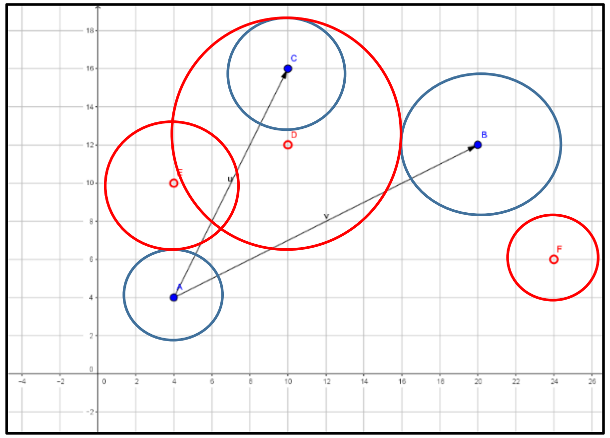
- **Region connected calculus (RCC8/RCC5) calculus**
  - 8 binary JEPD relations
  - Expresses relations between regions
  - Represents static or dynamic relations



Qualitative Spatial Representations

- **Region connected calculus (RCC8/RCC5) calculus**

Simple model:

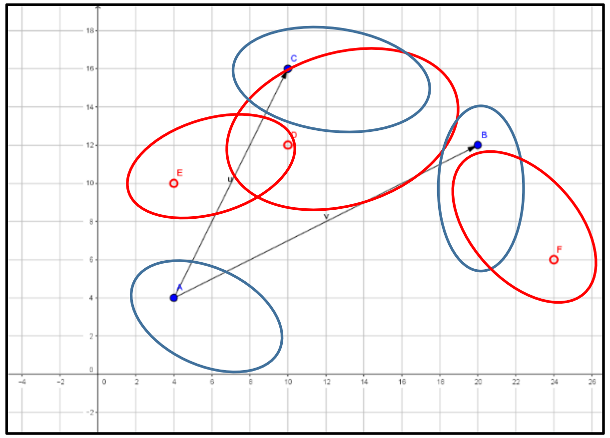


	A	B	C	D	E	F
A		DC	DC	DC	EC	DC
B	DC		DC	EC	DC	DC
C	DC	DC		TPP	DC	DC
D	DC	EC	TPI		PO	DC
E	EC	DC	DC	PO		DC
F	DC	DC	DC	DC	DC	

passer

actual receiver

Complex model:



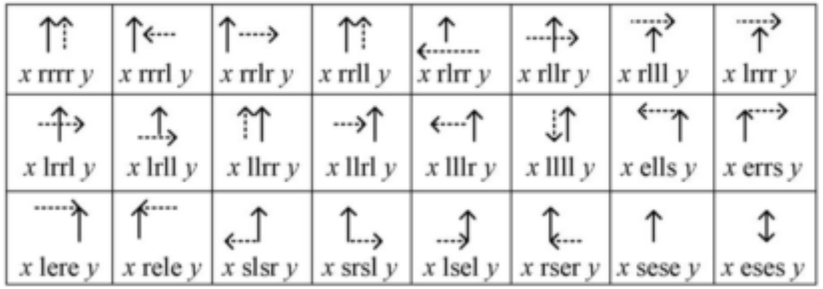
	A	B	C	D	E	F
A		DC	DC	DC	DC	DC
B	DC		DC	EC	DC	PO
C	DC	DC		PO	DC	DC
D	DC	EC	PO		PO	DC
E	DC	DC	DC	PO		DC
F	DC	PO	DC	DC	DC	

passer

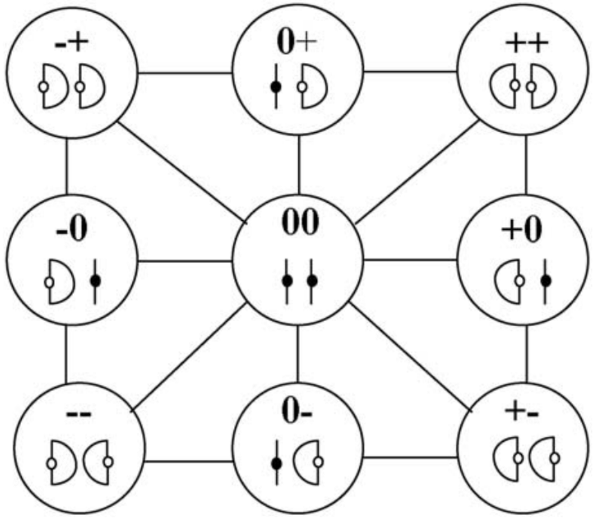
actual receiver

Qualitative Spatial Representations

- **Dipole calculus** OR **qualitative trajectory calculus**
  - Captures movement information
  - Both *spatial* and *temporal* information



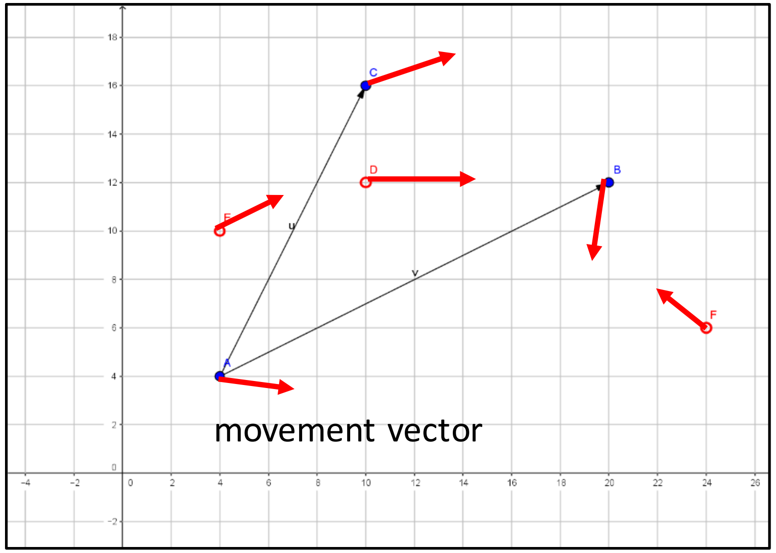
Dipole calculus



Qualitative trajectory calculus

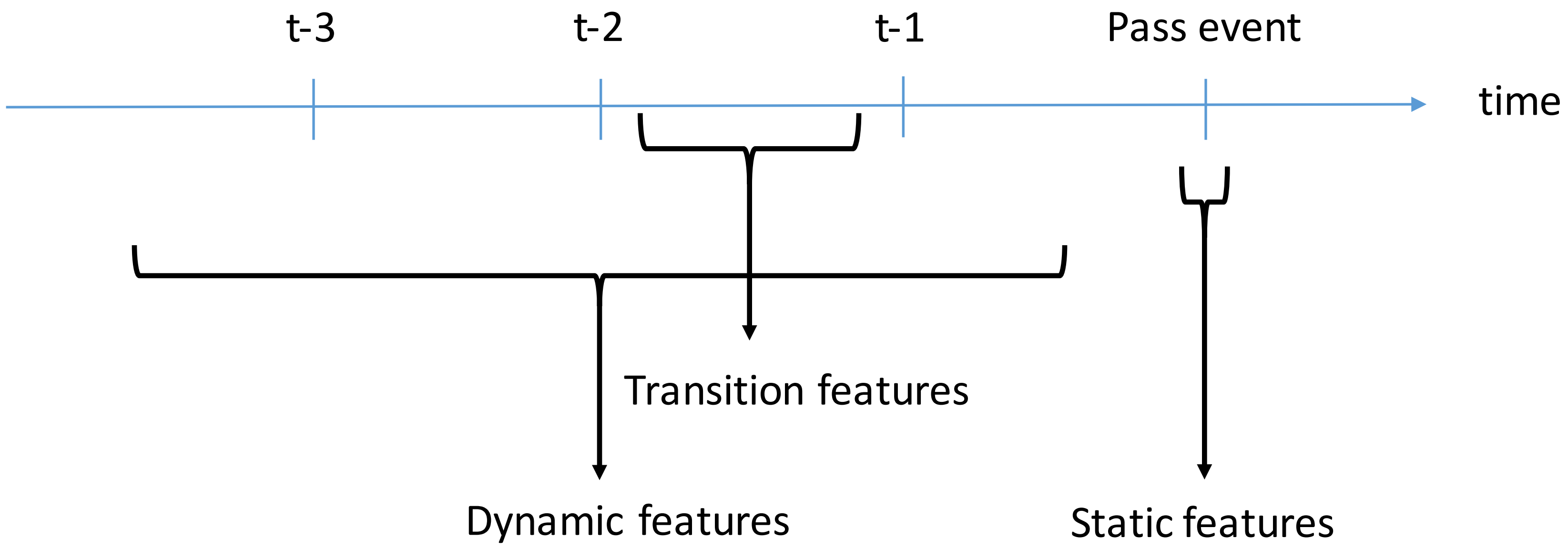
Qualitative Spatial Representations

- **Dipole calculus** OR **qualitative trajectory calculus**
  - Captures movement information
  - Both *spatial* and *temporal* information



	A	B	C	D	E	F	
A		***llrl***	**llrr**	**Llrr**	**llrr**	**llll**	**passer**
B	-		*errs*	*rlll*	*errs*	*rele*	*receiver*
C	-	-		rrrr	rrrr	rrrl	
D	-	-	-		llrr	Llll	
E	-	-	-	-		rele	
F	-	-	-	-	-		

Capture the dynamics



- Static features only capture information at the moment of the pass
- Dynamic features capture information in moments leading up to the pass
- Transition features describe the transition between moments

Learn a prediction model with ILP

- ILP = *Inductive logic programming*

variable

clause $\underbrace{pass(A,\ yes)}_{head} \leftarrow \underbrace{receiver(A, free) \wedge state(A, running) \wedge \underbrace{direction(A, goal)}_{atom}}_{body}$

- ILP allows to encode knowledge with *logic programs*
- The above rule states

“If player A is free and running towards the goal, I will pass to him”

→Ideal to encode the qualitative relations from the QSR’s
→We can express background knowledge in the dataset

Learn a prediction model

- ILP algorithm 1: TILDE
  - Learns a decision tree
  - Divide-and-conquer
  - Transform tree to rule-set
  - PROBLEM: not robust to skewed data distribution & increasing amount of features

- ILP algorithm 2: ALEPH
  - Separate-and-conquer
  - Learns theory (= set of rules) that classifies examples
  - Starts from bottom-clauses that are refined and selected according to criteria
  - More robust to skewed distribution & increasing amount of features

→ We can use the learned rules that encode $pass$ or $no\ pass$ to predict unseen cases

Contents

Evaluation metric

- Best evaluation metric is a ranking between players
  - Award higher score if the model ranks the actual receiver higher
  - Example

Example	A	B	C	D	E	...	J
1	1	4	6	3	10		2
2	4	2	1	6	8		5

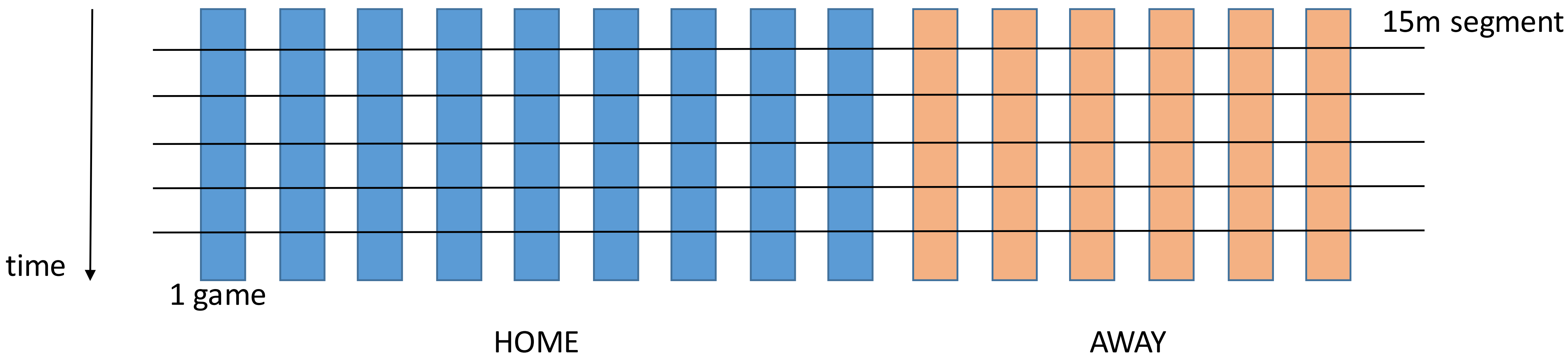
= actual receiver

  - Accuracy is only 0.5
  - Mean reciprocal rank (MRR) is 0.75

- Accuracy is a lower bound of the MRR:

$$Accuracy = \frac{\sum_{i=1}^{n} x_i}{n} \qquad \leq \qquad MRR = \frac{\sum_{i=1}^{n} \frac{1}{x_i}}{n}$$

Train and test data



- 14 games are available: 9 home and 6 away
- This allows us to construct some interesting sports-related hypotheses

Experimental hypotheses

- **Base hypothesis**:
  - Is the qualitative approach better than the quantitative at learning a meaningful model?
- **Sports-related questions**:
  - Is there a difference in the passing behaviour of a team at home and away?
  - Is there a decrease in performance throughout the game, altering passing behaviour?
  - Is passing behaviour team specific?

Results

1. Is the qualitative approach better than the quantitative at learning a meaningful model?

		MRR	top-1	top-2	top-3	Rules
Quant.	Non-rel.	0.11	0.84	0.93	0.93	8
	Rel.	0.24	10.82	18.16	21.76	524
Qual.	Static	0.39	25.49	36.33	41.22	582
	Dynamic	0.32	15.48	26.49	34.75	687
	Transition	0.33	17.48	29.24	35.00	681
	Combined	0.42	27.87	41.59	46.70	555

MRR = mean reciprocal rank

top-* = percentage of times the actual receiver is ranked accordingly by the learned model

Rules = number of logic rules in the learned theory

- A non-relational quantitative model cannot learn a meaningful model
- The qualitative approach is clearly better than a quantitative model
- The best model considers all information in the moments leading up to the pass

Results

1. Is there a difference in the passing behaviour of a team at home and away?
   → a home-trained model performs worse on away data and vice versa

2. Is athere a decrease in performance throughout the game, altering passing behaviour?
   → a model performs bets when it is applied to the same moment of the game it is trained on

3. Is passing behaviour team specific?
   → the model performs better when trained on a specific team and applied to that team

		MRR	top-1	top-2	top-3	Rules
1	Train home - test home	0.42	27.87	41.59	46.70	555
	Train home - test away	0.37	21.56	35.25	40.58	712
2	Train 1st half - test 1st half	0.42	27.87	41.59	46.70	555
	Train 1st half - test 2nd half	0.38	27.15	31.95	36.03	620
3	Train 1 team - test multiple	0.28	13.15	22.44	30.00	591
	Train multiple - test multiple	0.37	23.71	35.05	40.33	381

Contents

Conclusion

- Main messages:
  - qualitative, relational approach learns meaningful models
  - dynamics of the game are important
- Questions?