

Predicting Soccer Highlights From Spatio-temporal Match Event Streams

Tom Decroos

Vladimir Dzyuba

Jan Van Haaren

Jesse Davis

{firstname.lastname}@cs.kuleuven.be
http://dtai.cs.kuleuven.be/sports

Motivation: Highlight channels



Q: How do broadcasters choose which soccer match to display?

A: Broadcasters switch to a match **AFTER** an interesting event has happened in that match.

Research question:
Can we switch **BEFORE** an interesting event happens?

Problem statement

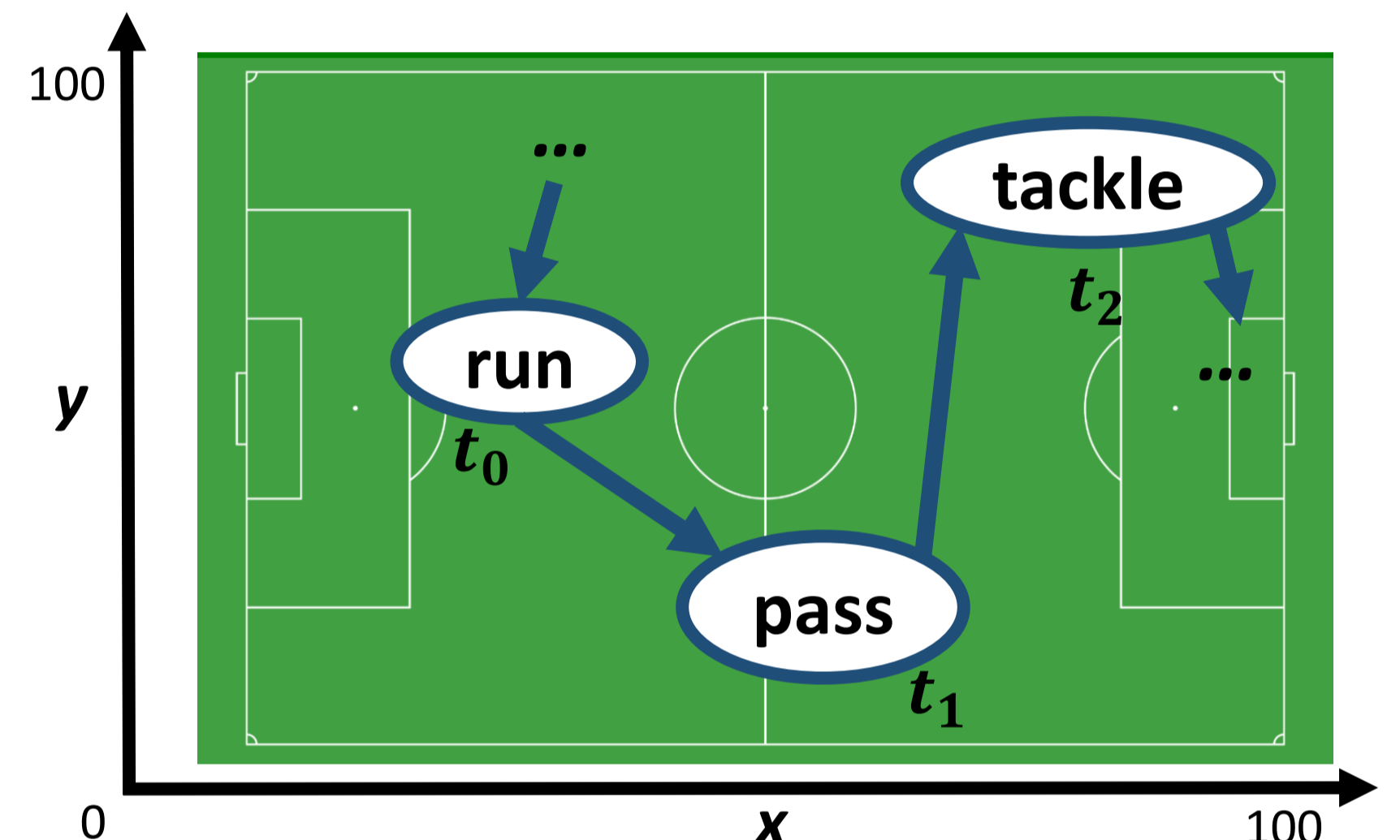
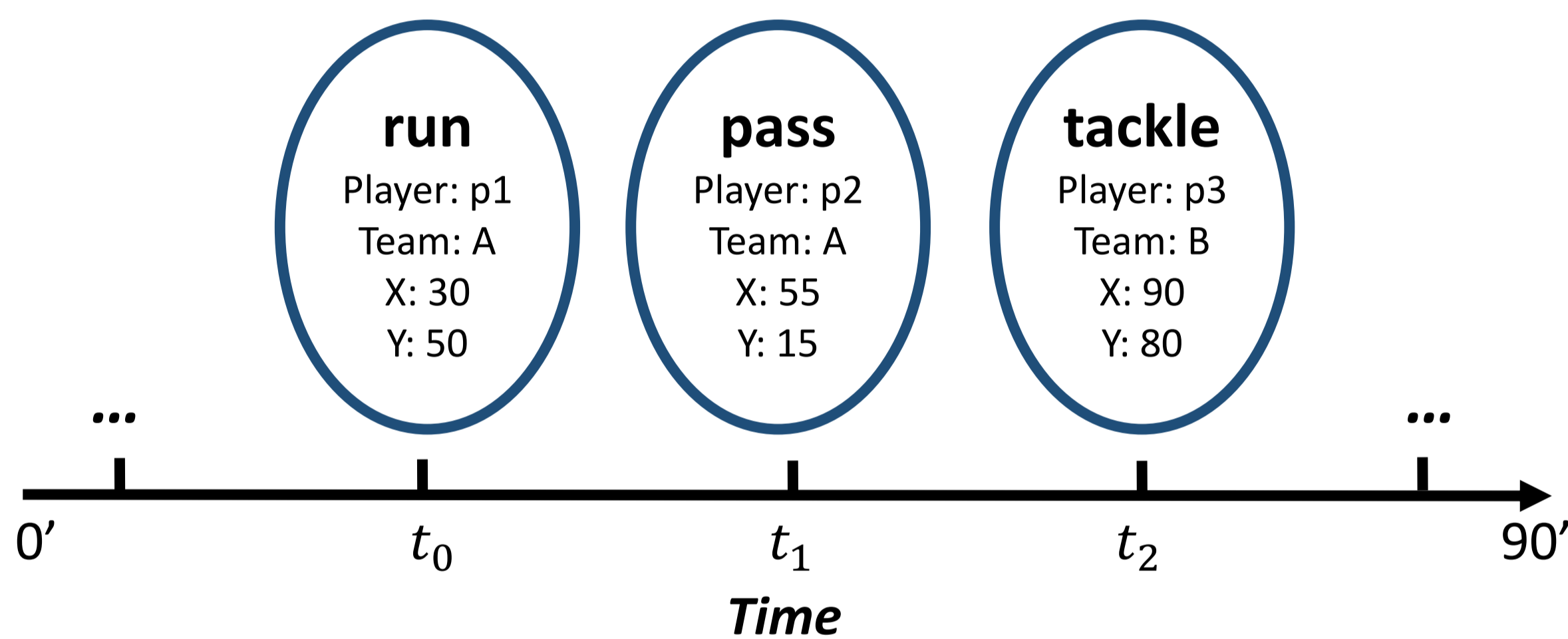
Given: A soccer match at time T .

Predict: The probability of a goal being scored in the next δ seconds.

Challenges

- Goals are rare. Not enough data to learn a conditional distribution.
- How to handle the spatio-temporal aspect of soccer gameplay?

Spatio-temporal match event streams



Our approach: POGBA (Prediction Of Goals By Assessing phases)

Key ideas

- Goals are almost always preceded by shots.
- Use a generative model to estimate goal probability.

$$P(\text{goal}, \text{shot}, \text{phase}) = P(\text{goal}|\text{shot}) \times P(\text{shot}|\text{phase}) \times P(\text{phase})$$

Goal probability depends mostly on shot quality.

Shots are more frequent than goals.

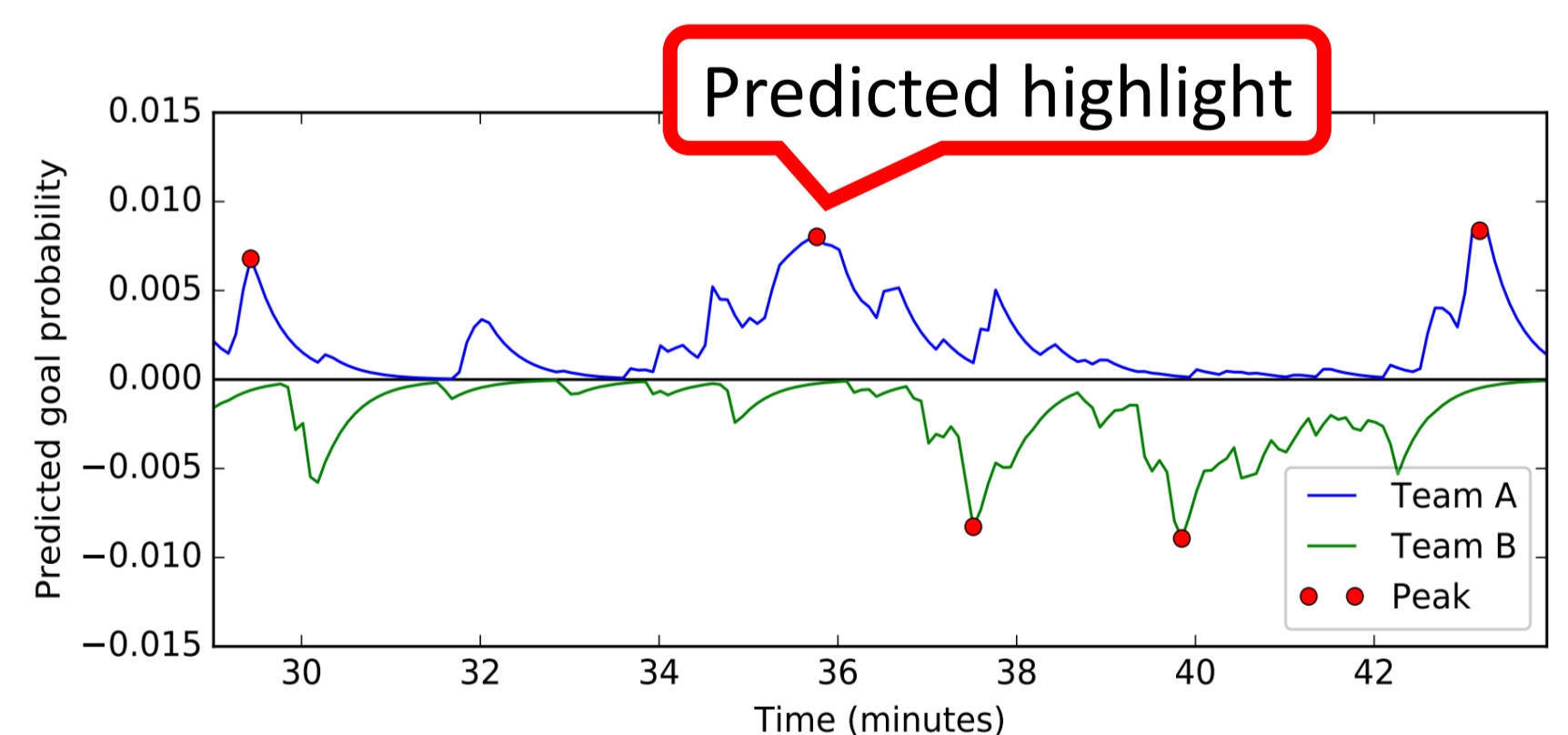
$$P(\text{goal}|\text{phase}) = \sum_{\text{shot}} P(\text{goal}|\text{shot}) \times P(\text{shot}|\text{phase})$$

Expected Goals Model

Spatio-temporal Nearest Neighbours

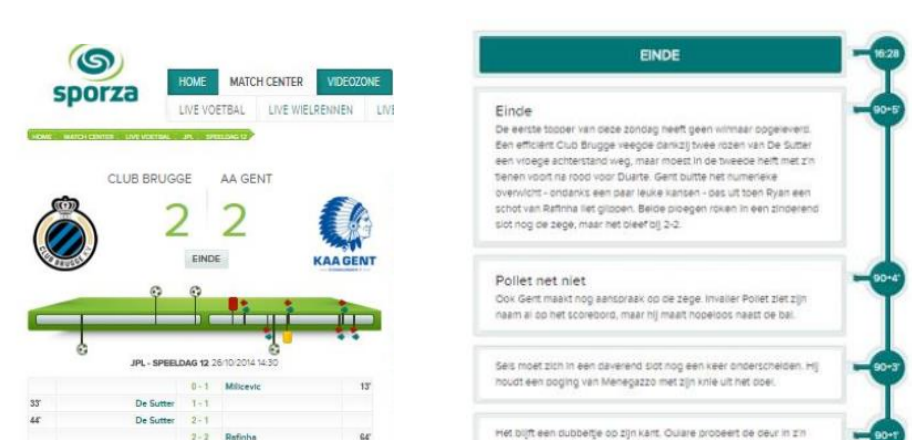
Predicting highlights using POGBA

Interesting events: high probability that a goal will be scored in the near future.



Experiments

- Event streams of 69 matches provided by professional soccer club.
- Highlights of 25 matches from sports website used for evaluation.



| | Temporal aspect | Indirect estimation | Precision | Recall | F_1 |
|--------------|-----------------|---------------------|-----------|--------|-------|
| POGBA | ✓ | ✓ | 0.44 | 0.61 | 0.51 |
| SpatTempDir | ✓ | — | 0.43 | 0.59 | 0.50 |
| SpatIndir | — | ✓ | 0.36 | 0.46 | 0.40 |
| SpatDir | — | — | 0.34 | 0.39 | 0.36 |
| Final 4th | — | — | 0.22 | 0.72 | 0.33 |
| Random | — | — | 0.05 | 0.05 | 0.05 |

Conclusions

- Handling the spatio-temporal aspect of the data is essential.
- POGBA can also be applied to other domains with important events (industrial machines, hospitals, ...).
- POGBA can generate advanced stats for soccer analytics.