Deep Learning from Spatial Relations
for Soccer Pass Prediction

Ondřej Hubáček, Gustav Šourek, and Filip Železný

Czech Technical University, Prague, Czech Republic
{hubacon2,souregus,zelezny}@fel.cvut.cz

Abstract. We propose a convolutional architecture for learning representations over spatial relations in the game of soccer, with the goal to predict individual passes between players, as a submission to the prediction challenge organized for the 5th Workshop on Machine Learning and Data Mining for Sports Analytics. The goal of the challenge was to predict the receiver of a pass given location of the sender and all other players. From each soccer situation, we extract spatial relations between the players and a few key locations on the field, which are then hierarchically aggregated within the neural architecture designed to extract possibly complex gameplay patterns stemming from these simple relations. The use of convolutions then allows to efficiently capture the various regularities that are inherent to the game. In the experiments, we show very promising performance of the method.

1 Introduction

Predictive sport analytics is a modern discipline where various statistical models are employed to assess different aspects of a game. In this paper, we focus on the game of soccer and predicting individual passes between players during the match, given a static snapshot of each pass situation, i.e. indication of ball possession and locations of all the players. This setting was given by the prediction challenge organized for the 5th Workshop on Machine Learning and Data Mining for Sports Analytics held in conjunction with ECML.

Since each learning example is, in this case, just an independent, static viewpoint on the game, we approach the problem from a simple geometrical perspective. In that view, we take each situation, determined by mere absolute locations of the players, enrich these with a few soccer-specific contextual locations, and turn their absolute positions into mutual, relative distances. This way we enable to the model to generalize across different situations, reasoning about the mutual spatial patterns between the players, rather than their positions on the filed. These spatial patterns are represented with convolutional filters, capturing the inherent symmetries and geometrical regularities arising from the rules of the game. These patterns are then further aggregated with pooling and combined in a fully connected manner to help the model to explore their relations. As opposed to some existing works based quite heavily on expert knowledge, we employ just a very few assumptions on the patterns and rather aim at the benefits of end-to-end learning.

1.1   Related Work

Inductive logic programming model [7] trained on qualitative spatial representations [2] was previously used to tackle the task of predicting soccer passes. Similar approach was used for discovering offensive patterns [6]. Spatio-temporal data were further utilized to infer teams' play-styles [1,4] and to examine the likelihood of scoring a goal from a shot [3]. Another approach leveraged a physics-based model of soccer ball motion to predict the receiver of the pass [5].

1.2   Dataset

The dataset consisted of 12 124 soccer passes from which 10 045 passes were successful (meaning that the sender and the receiver of the pass were from the same team). We decided to focus on predicting only the successful passes as was done previously [7].

Unlike in the previous work [7], the dataset contained solely the snapshot of the game in form of coordinates of each of the 22 players on the field, making the situations independent of each other. This makes the prediction task much harder, because we have no information about players' momentum, or orientation in space etc. Neither were are capable to determine the same team or player across multiple situations. The dataset also contained the timestamp of when the pass was send and received. Due to the predictive nature of the task, we decided to omit the timestamp of the pass receipt, since it is obviously not available when making the actual prediction.

In 367 cases, only 21 players' coordinates were present, presumably after one player had been sent off. To deal with the missing coordinates we inputted surrogate large numbers as the coordinates, so this position became meaningless for the predictions.

2   Predictive model

The proposed model is a neural architecture consisting of a convolutional layers with diverse filters, max-pooling and a fully connected layers with a softmax output. Each of the convolutional filters encodes a certain feature-set transformation designed to extract a particular context from the game snapshots. Intuitively, these may collect information on how occupied the potentially receiving player is, how pressured the sender of the pass is, or where is the receiver positioned on the field w.r.t. his teammates. The max-pooling layer helps the model to become agnostic to the particular positioning and ordering of the players in order to generalize better, based on the intuition that typically only a very few closest players are relevant to each pass. The softmax output then naturally encodes the exclusive outcomes of each situation, since only one pass at a time is ever carried out.

Table 1. Enriching spatial snapshots with contextual locations.

Feature	Description
$f_1 : dist(p_s, p_r)$	Distance between sender p_s and potential receiver p_r.
$f_2 : dist(p_s, goal)$	Distance of p_s to the center of the opponent's goal.
$f_3 : dist(p_s, side)$	Perpendicular distance of p_s to closest sideline.
$f_4 : dist(p_s, center_{t_1})$	Distance of p_s to the center of gravity of his teammates.
$f_5 : dist(p_s, center_{t_2})$	Distance of p_s to the center of gravity of his opponents.
$f_6 : dist(p_r, goal)$	Distance of p_r to the center of the opponent's goal.
$f_7 : dist(p_r, side)$	Perpendicular distance of p_r to closest sideline.
$f_8 : dist(p_r, center_{t_1})$	Distance of p_r to the center of gravity of his teammates.
$f_9 : dist(p_r, center_{t_2})$	Distance of p_r to the center of gravity of his opponents.
$f_{10} : dist(p_r, p_i)$	Distance of p_r to a teammate p_i.
$f_{11} : dist(p_s, p_i)$	Distance of p_s to a teammate p_i.
$f_{12} : dist\_opp(p_r, p_i)$	Distance of p_r to an opponent p_i.
$f_{13} : dist\_opp(p_s, p_i)$	Distance of p_s to an opponent p_i.

2.1 Knowledge Representation

The raw data come in a simple table format where, for each pass situation during the course of each game, we are given x-y coordinates of the 22 players on the field with an indicator of the sender of the pass p_s, i.e. a tuple of

$$(timestamp, p_{1_x}, p_{1_y}, \ldots, p_{11_x}, p_{11_y}, p_{12_x}, p_{12_y} \ldots, p_{22_x}, p_{22_y}, p_s)$$

For the purpose of pass prediction, we look at each snapshot from the perspective of potential successful passes between the ball-possesing player p_s and all his teammates (potential receivers) p_r, i.e. for each situation we have 10 pairs of players

$$(p_s, p_r), \text{ such that } p_r \in \begin{cases} \{p_1, \ldots, p_{11}\} \setminus p_s, & \text{if } s \in \{1, \ldots, 11\} \\ \{p_{12}, \ldots, p_{22}\} \setminus p_s, & \text{if } s \in \{12, \ldots, 22\} \end{cases}$$

As a preprocessing step, we enrich these pairs with several key static and dynamic locations from the field, upon which we measure distances as described in Table 1. These enriched pairs, representing the potential passes, then constitute our learning examples.

2.2 Neural Architecture

An overview of the neural architecture is displayed in Figure 1. At the input to the model, the resulting spatial relations described in Section 2.1 are being aggregated into sets to form feature maps for the convolutional filters. Particularly, for each potential pass (p_s, p_r), we conform the relations into different filters expressing different viewpoints on the pass, such as *cover* of the receiving player or *pressure* on the sender and *alternatives* available to him, as detailed in Table 2.

**Table 2.** Conformation of spatial relations into convolutional filters.

Filter	Features	Context
alternative	(f_1, f_{10}, f_{11})	The alternatives available to the sender.
cover	(f_1, f_{12}, f_{13})	The occupation of the sender by his opponents.
pressure	(f_1, f_{12}, f_{13})	The pressure on the sender from his opponents.

Each of these feature sets, or filters, may be instantiated multiple times w.r.t. the variables p_i iterating over the opponents of the sender (*cover*, *pressure*) and teammates of the sender (*alternative*). Within the context of each filter, we order the remaining players w.r.t. the f_{10}, f_{12} and f_{13} for *alternative*, *cover* and *pressure*, respectively. This way we enforce ordering on these instantiations, resulting into 1D feature maps upon which the filters operate, as depicted in Figure 1. Thus despite the *cover* and *pressure* filters operating on the same feature sets, they will result into different feature maps. Also, since all these filters principally share the common static context of where within the field the current situation occurs, described by the features $f_1 \ldots f_9$, we exclude these from the individual filters to merge them later in the model only to prevent redundancy in the feature maps.

The resulting values from these filters are then aggregated via max-pooling. While multiple pools could be connected with a standard overlay to capture the different sub-regions of the distance space, we set a global pool over all instantations of each single filter, following the intuition that only the closest players are typically relevant, suppressing the potential noise from the rest. To alleviate this somewhat radical assumption, we also employ wider filters to capture couples of the remaining players rather than individuals. This way we may also reason about more complex spatial patterns between the relevant players. These filters of size 3×2 and 3×3 further distinguish the use of *cover* and *pressure*. Finally the pooling helps to neglect the potentially harmful effect of the, to a certain degree ad-hoc, overall ordering.

The patterns extracted with the help of the filters and selected by the pools form an input to the fully connected layers (Figure 1). The purpose of these layers is to combine all the different patterns into a final value expressing the potential of each individual pass (p_s, p_r). Intuitively, these layers express the logic of decision making the sender p_s is normally going through, incorporating the relational contexts (filters) of the receiving player p_r w.r.t. his own, while learning how to weight the importance of the individual patterns in each combination.

Finally, with the softmax output (Figure 1), we enable the model to reason jointly over the whole set of 10 possible passes (p_s, p_r). As opposed to separating each pass situation into 10 independent learning examples and normalizing over these as a postprocessing step, with the joint output the gradient directly steers the model towards exclusive predictions as part of the learning process.
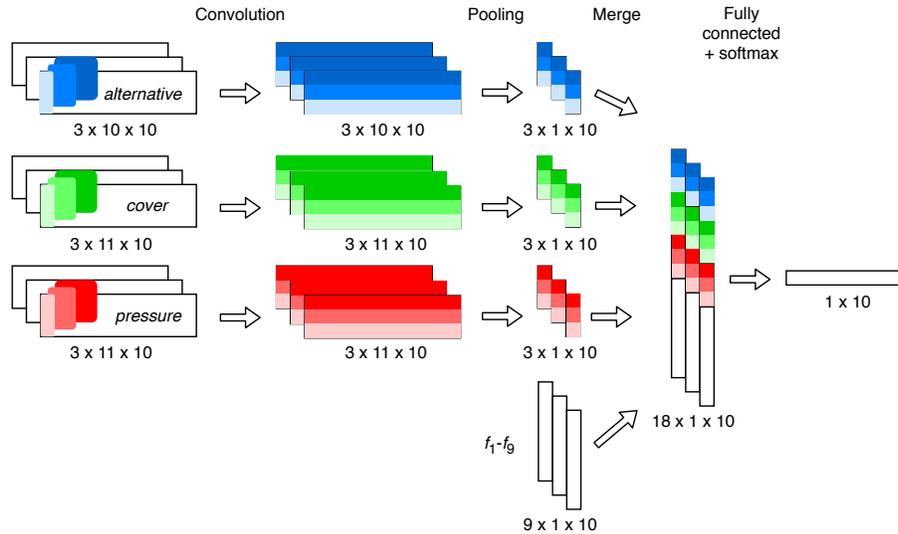
Fig. 1. Architecture of the neural model. Four feature maps of size $\#features \times \#instantiations \times \#possibilities$ are at the input. Filters of size 3×1, 3×2 and 3×3 are applied to each feature map. The outputs of the convolution are reduced by max pooling and merged with the $f_1 - f_9$ features providing their static context. Finally, 2 dense layers with 3, respectively 1, neurons are applied to each possibility. For clarity only 3 out of 10 possibilities (depth dimension) are displayed.

3   Experiments

We performed 10-fold crossvalidation, evaluated the model w.r.t. mean reciprocal rank and how many times the actual receiver of the pass was among the three most likely predictions. We compared our result with [7], where the authors made use of both static and dynamic features derived from the flow of the game, which was unavailable to us. Therefore it would be fair to compare our results with the *Static* model from the mentioned work. Nevertheless, our model outperformed both the *Static* and the *Combined* model, which combined static and dynamic features (Table 3).

3.1   Human-level Performance

We measured human-level performance to assess the inherent difficulty of the task. We were particularly curious about the effect of the missing dynamic context of the game that humans are used to from standard visual recordings, providing much more information than the mere static snapshots. We measured and averaged the predictive performance of three soccer enthusiasts on a sample

Table 3. Comparison of the model's (CNN) performance with previous work [7] and human-level performance.

	MRR	top-1	top-2	top-3
Static	0.39	25.49	36.33	41.22
Combined	0.42	27.87	41.59	46.70
CNN	0.48	29.87	45.63	55.91
Human	0.56	34.00	56.25	70.25

of 200 randomly selected situations. To put the data into a more familiar perspective, we created a simple interactive visualization[1] that may be utilized for further measurements. The task proved to be difficult even for humans. While the top-1 accuracies of the model and humans were close, the top-3 accuracies and MMR showed that humans clearly rank the alternatives better.

3.2   Discussion

We analyzed the predictions made by the model to obtain further insights. The main weakness of the model was that it usually considered only a few options as viable, even when their alternatives were very similar. This could be due to the use of softmax in combination with cross-entropy loss when training the network instead of some kind of ranking loss. The network was strong in spotting uncovered teammates, sometimes even overvaluing their positions. Generally, the network preferred passes to sidelines, even when we could guess that the ball most likely just came from those positions. The human intuition thus seems superior when capturing this underlying "flow" of the game.

Vizualization of an example situation illustrates the difficulty of the task 2. Without the information about the senders orientation on the field, there are many viable alternatives. The model marked the pass to the sideline as the most probable as this is a common pattern – midfielder developing a play from the center of the field with a pass to the sideline. The actual pass was the model's second guess. From a human perspective there are far too many options assigned near zero probability. Especially the passes to the players 9 and 2 should have been prioritized more.

The decomposition of the static context features $f_1 \ldots f_9$ from the convolutional filters, as depicted in Figure 1, might suggests that the model could be split into two. While these complementary feature sets provide context to each other and were thus meant to work together, we also measured their separate performance, proving the convolutional features to be more valuable (MRR 0.46) than the static context features (MRR 0.42) in separate experiments.

---

[1] https://github.com/Hudler/pass-viz
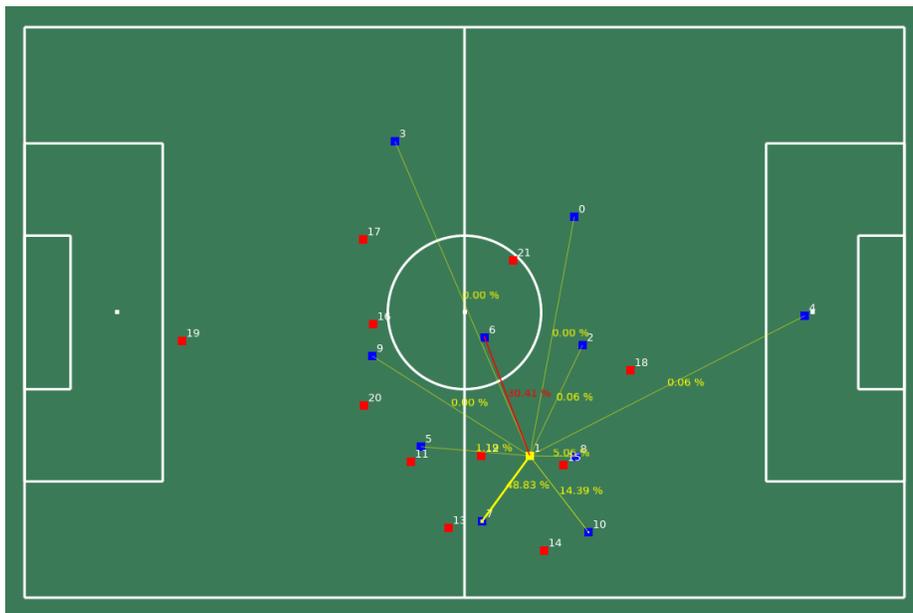
Fig. 2. Example model prediction. Possible passlines are depicted by yellow lines, with the actual pass marked by red. The percentages near the passlines show the predicted probabilities.

4   Conclusion

We detailed our model for soccer pass prediction given static spatial snapshots of the game. The model was a neural architecture based on a set of convolutional filters, carefully designed to extract different relational contexts from each game situation, i.e. mutual positions of players on the field. We argued how such an architecture may learn possibly complex relational patterns via aggregation of simple spatial relations. Finally, on a large dataset of captured soccer passes, we showed that promising results can be achieved with such an approach.

References

1. Joel Brooks, Matthew Kerr, and John Guttag. Using machine learning to draw inferences from pass location data in soccer. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, 9(5):338–349, 2016.
2. Juan Chen, Anthony G Cohn, Dayou Liu, Shengsheng Wang, Jihong Ouyang, and Qiangyuan Yu. A survey of qualitative spatial representations. *The Knowledge Engineering Review*, 30(1):106–136, 2015.
3. Patrick Lucey, Alina Bialkowski, Mathew Monfort, Peter Carr, and Iain Matthews. quality vs quantity: Improved shot prediction in soccer using strategic features from spatiotemporal data. In *Proc. 8th annual mit sloan sports analytics conference*, pages 1–9, 2014.
4. Patrick Lucey, Dean Oliver, Peter Carr, Joe Roth, and Iain Matthews. Assessing team strategy using spatiotemporal data. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1366–1374. ACM, 2013.
5. William Spearman, Austin Basye, Greg Dick, Ryan Hotovy, and Paul Pop. Physics-based modeling of pass probabilities in soccer. In *Proceeding of the 11th MIT Sloan Sports Analytics Conference*, 2017.
6. Jan Van Haaren, Vladimir Dzyuba, Siebe Hannosset, and Jesse Davis. Automatically discovering offensive patterns in soccer match data. In *International Symposium on Intelligent Data Analysis*, pages 286–297. Springer, 2015.
7. Vincent Vercruyssen, Luc De Raedt, and Jesse Davis. Qualitative spatial reasoning for soccer pass prediction. In *Machine Learning and Data Mining for Sports Analytics ECML/PKDD 2016 workshop*, 2016.