# Spatio-Temporal Convolution Kernels for Clustering Trajectories

Konstantin Knauf
Knowledge Mining and Assessment
Technische Universität Darmstadt
Darmstadt, Germany
knauf.konstantin@gmail.com

Ulf Brefeld*
Knowledge Mining and Assessment
Technische Universität Darmstadt
Darmstadt, Germany
brefeld@kma.informatik.tu-darmstadt.de

## ABSTRACT

We propose a novel class of kernels to identify tactical patterns in multi-trajectory data such as soccer games. Formally, we introduce a group of $R$-convolution kernels called Spatio-Temporal Convolution Kernels composed of a temporal and a spatial kernel. The particular choice of the component kernels depends on the application at hand. For the purpose of clustering player and ball trajectories in soccer we propose a probability product kernel on the empirical distributions of the objects to serve as spatial kernel and a Gaussian kernel as temporal kernel. Empirically, we observe better clusterings compared to baseline methods and high cluster consistencies with (inefficient) Dynamic Time Warping-based methods. In terms of tactical patterns we identify interpretable clusters corresponding to long and short game initiations on either sides of the field.

## 1. INTRODUCTION

Tracking moving objects is a prerequisite for analysing coordination and drawing conclusions on optimality, strategy, and tactics. Applications are not restricted to sports analytics but also include video surveillance, animal migration and traffic analysis. In all these applications the movement of (interesting) groups of objects is more informative than the trajectory of a single object.

By contrast, existing approaches on trajectory analyses often focus on the analysis of trajectories from single objects [3, 7, 9, 13, 14]. Junejo et al. [9] represent trajectories as a set of two-dimensional coordinates. Using Hausdorff distance and graph-cuts trajectories are then recursively partitioned. Fu et al. [3] resample trajectories to obtain constant between-point distances. The corresponding points of two trajectories are compared using a Gaussian RBF Kernel, whereby the longer trajectory is cut to the length of the shorter one. Hirano et al. [7] use multi-scale matching together with a rough clustering to analyse trajectories. Grimson et al. [13] represent trajectories by a bag-of-positions/directions similar to the bag-of-words representation of documents in natural language processing. The spatial domain is discretized and the number of occurrences of each position/direction in a trajectory is counted. A semantic topic model called Dual-HDP is used to find semantic regions that are considered building blocks of trajectories. Wei et al. [14] use role models and a bilinear spatio-temporal basis model to represent team movement to cluster goal scoring opportunities in soccer.

We propose multi-purpose kernels to represent, find and compare groups of related trajectories. Representing trajectories as a set of time/positions-tuples, we employ an $R$-convolution kernel with two main components: a spatial component comparing snapshots of an arbitrary number of objects and a temporal component introducing a temporal order on these snapshots. Empirically, we evaluate our approach on team tactics in soccer. Note that this is a particularly challenging task as the data is noisy and often unstructured due to the continuous nature of the game and individual short-term goals of the players. For lack of space, we focus on game initiations, however, similar results are obtained for scoring opportunities.

The remainder of this paper is structured as follows. Section 2 describes our method by deriving a so called Spatio-Temporal Convolution Kernel. In Section 3 we deploy our method to identify tactical patterns in a soccer game and Section 4 concludes.

## 2. PROPOSED APPROACH

### 2.1 Trajectory Representation

Multi-object trajectory analysis deals with a possibly varying number of moving objects $\mathcal{O}_t$ in a metric space $X$, e.g. $X = \mathbb{R}^2$, over a period of time $\mathcal{T} \subset \mathbb{R}^+$. A multi-object trajectory is composed of snapshots of the positions of the objects at multiple times, defined by Definition 1.

DEFINITION 1. *Assume there is a constant number of groups $K$, such that at any time every object can be associated with one of the groups $k \in \{1, ..., K\}$. Then the* ***group-oriented snapshot*** *of all objects at time $t$ is denoted by*

$$x_t \in \mathcal{P}(X)^K =: \mathcal{X}.$$

*We will call $\mathcal{X}$ the* ***snapshot space****. The positions of all objects of a particular group $k \in \{1, ..., K\}$ is denoted by*

$$x_t(k) \in \mathcal{P}(X),$$

*UB is also affiliated with the German Institute for Educational Research (DIPF), Frankfurt/Main, Germany


*2014 KDD Workshop on Large-Scale Sports Analytics* '14 New York, New York USA

*where the members of group $k$ in snapshot $x_t$ are denoted by $O_{x_t}(k) \subset \mathcal{O}_t$.*

Instead of an ordered sequence of snapshots we will use a set of time/snapshot-tuples to represent trajectories. Thus, time is explicitly represented in contrast to the implicit representation as sequences. Using Definition 1, trajectories are defined by Definition 2.

DEFINITION 2. *A **trajectory** is defined as a finite subset $P = \{t_1, x_{t_1}), ..., (t_n, x_{t_n})\} \subset [0,1] \times \mathcal{X}$, such that the trajectory set $P$ contains only one snapshot per point in time and the time-scale is normalized to $[0,1]$.*

## 2.2 Spatio-Temporal Convolution Kernel

In this section we develop a kernel on the space of (time-normalized) multi-trajectories $\mathcal{P}([0,1] \times \mathcal{X})$. Each of the trajectories consists of a set of snapshots associated with a relative time stamp. The basic idea is to perform a pairwise comparison of the snapshots in the two sets. Therefore, first, we need a way to compare snapshots and, second, we need to know, which snapshots of the two trajectory sets to compare with each other. At first sight, Dynamic Time Warping (DTW) [1] seems to be a suitable candidate as it computes the best alignment of the snapshots. However, besides the high computational costs of DTW, the obtained similarity measure is not a Mercer kernel, i.e. does not correspond to an inner product in some Hilbert Space. Although there is anecdotal evidence that learning with indefinite kernels can lead to good results in some application, theoretical guarantees only exist for positive definite kernels.

We propose to compare every snapshot of the first trajectory to every snapshot of the second one and to weight the similarities according to their offset in relative time. Formally, this is done using an $R$-convolution kernel [6] on the two sets representing the trajectories. In order to use an $R$-convolution kernel we need to define a function $R$,

$$R : \mathbb{N} \times [0,1] \times \mathcal{X} \times \mathcal{P}([0,1] \times \mathcal{X}) \to \{0,1\},$$

relating the trajectory sets to their components:

$$R(n,t,x,P) = \begin{cases} 1 & \text{if } |P| = n \wedge \exists (s,y_s) \in P : (t,x) = (s,y_s) \\ 0 & \text{otherwise} \end{cases}$$

The $R$-convolution kernel is then given by

$$k(P,Q) = \sum_{\substack{(n,t,x) \in R^{-1}(P), \\ (m,s,y) \in R^{-1}(Q)}} k_{\mathbb{N}}(n,m) \cdot k_{[0,1]}(t,s) \cdot k_{\mathcal{X}}(x,y) \quad (1)$$

with $R^{-1}(P) = \{(n,t,x) : R(n,t,x,P) = 1\}$. The term $k_{\mathbb{N}}$ accounts for differences in the length of trajectories by normalization, i.e. $k_{\mathbb{N}} = \frac{1}{nm}$. Finally, the $R$-convolution kernel simplifies to

$$k(P,Q) = \frac{1}{|P||Q|} \sum_{(t,x_t) \in P, (s,y_s) \in Q} k_{[0,1]}(t,s) \cdot k_{\mathcal{X}}(x_t, y_s) \quad (2)$$

The definition of the Spatio-Temporal Convolution Kernel (STCK) in Equation (2) leaves two degrees of freedom. First, the definition of the spatial kernel $k_{\mathcal{X}}$ that determines when snapshots are similar. Second, the choice of the temporal kernel $k_{[0,1]}$ that determines the way in which the snapshots of two sequences are combined and thus the importance of ordering and speed.

The actual choice of both kernels depends on the application at hand. In our case of soccer analytics, we propose to compare the $K$ groups separately for the **spatial kernel** and to sum up the individual contributions

$$k_{\mathcal{X}}(x_t, y_t) = \frac{1}{K} \sum_{k=1}^{K} k_G(x_t(k), y_t(k)). \quad (3)$$

According to Definition 1, kernel $k_G$ needs to compare two sets of a possibly varying number of positions. This could be done straight forwardly using a Gaussian RBF Kernel (or any other kernel on $\mathbb{R}^2$) on the centroids of the positions. Besides the need of defining the width parameter $\sigma_S$ this kernel has the drawback, that the distribution of the objects around their centroid is not taken into account. We deal with both problems by using a probability product kernel [8] on the Gaussian distributions, which are fitted to the positions of the objects of the two groups. We denote these distributions, respectively their density functions, by $\mathcal{N}(\mu_{x_t}(k), \Sigma_{x_t}(k))$ and $\mathcal{N}(\mu_{y_t}(k), \Sigma_{y_t}(k))$, respectively $g_{x_t}(k)$ and $g_{y_t}(k)$. The probability product kernel (with $\rho = 1/2$) between two Gaussian distributions provides a closed-form in terms of the means and covariance matrices and is given by

$$k_G(x_t, y_t) = \int_{\mathbb{R}^2} (g_{x_t}(z) g_{y_t}(z))^{1/2} \, \mathrm{d}z = 2|\Sigma^*|^{\frac{1}{2}} |\Sigma_{x_t}|^{-\frac{1}{4}} |\Sigma_{y_t}|^{-\frac{1}{4}}$$

$$\cdot \exp\left(-\frac{1}{4}\left(\mu_{x_t}^T \Sigma_{x_t}^{-1} \mu_{x_t} + \mu_{y_t}^T \Sigma_{y_t}^{-1} \mu_{y_t} - \mu^{*T} \Sigma^* \mu^*\right)\right)$$

with $\Sigma^* = (\Sigma_{x_t}^{-1} + \Sigma_{y_t}^{-1})^{-1}$ and $\mu^* = \Sigma_{x_t}^{-1} \mu_{x_t} + \Sigma_{y_t}^{-1} \mu_{y_t}$ [1]. If the covariance matrix is ill-conditioned or singular, we use a simple shrinking scheme with shrinkage parameter $\alpha$ to achieve non-singularity

$$\Sigma = (1-\alpha) \cdot \Sigma + \alpha \cdot \frac{\mathrm{Tr}(\Sigma)}{2} \mathbb{I}_2.$$

There exist different strategies to choose an optimal value for $\alpha$ (see [2, 11]), but for our purposes it suffices to deploy a constant 0.1. In case of $\mathrm{Tr}(\Sigma) = 0$ the following scheme is used:

$$\Sigma = (1-\alpha) \cdot \Sigma + \alpha \cdot \sigma_{MIN}^2 \cdot \mathbb{I}_2,$$

with an application specific parameter $\sigma_{MIN}^2$.

As for the **temporal kernels** the situation is much easier, because the space is fixed to the one-dimensional interval $[0,1]$. We will briefly discuss possible options for the temporal kernel and their implications.

- Constant Kernel $k_{[0,1]}(t,s) = 1$: If a constant kernel is applied the Spatio-Temporal Convolution Kernel collapses to a set kernel on the two sets of snapshots ignoring order at all.
- Uniform Kernel $k_{[0,1]}(t,s) = \mathbb{I}_{\{|t-s|<w\}}$: If a uniform kernel is used every snapshot of the first trajectory is just compared to those snapshots of the second trajectory, which are close in time. The choice of $w$ determines how close snapshots have to be.
- Gaussian Kernel $k_{[0,1]}(t,s) = \exp\left(-\frac{1}{2\sigma_T^2}|t-s|^2\right)$: In case of a Gaussian Kernel every snapshot of the first

[1] In order to simplify the notation the index $k$ has been omitted.

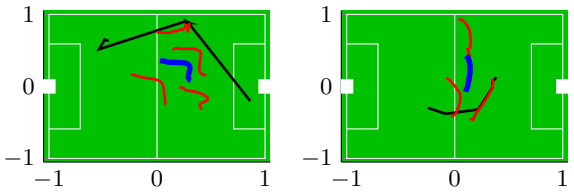


Figure 1: Exemplary game initiations: The trajectory of the ball (black), the four players of interest (red) and their centroid (blue) are depicted.

trajectory is compared to every snapshot of the second trajectory, but the closer they are in (relative) time, the more important their similarity is.

Based on our experiments with artificial data we propose a Gaussian kernel as temporal kernel.

## 2.3 Kernelized K-Medoids

For clustering we use a kernelized version of the K-Medoids algorithm [10], which has consistently outperformed spectral clustering in our tests. Similar to K-Means, K-Medoids is a partitional clustering algorithm, which aims to minimize the distance between the instances and the cluster centers they are associated to. K-Medoids chooses only instances itself as cluster centers and therefore only needs the distances between all instances as input, which are derived from the kernel using

$$dist(x, y) = k(x, x) - 2k(x, y) + k(y, y) \quad (4)$$

Medoids are then initialized randomly and an iterative optimization procedure similar to the K-Means algorithm is used to find a local minimum. This process is repeated 200 times and the clustering with the lowest within-cluster distance serves as the final clustering.

# 3. EXPERIMENTAL RESULTS

The goal of this section is to identify movements patterns in a soccer match through the analysis of tracking data. In particular, we analyse tracking data from the soccer match between 1. FC Kaiserslautern and Hannover 96 from the 2011/12 Bundesliga Season (Matchday 17). The data consists of two dimensional positions of players, ball and referees at 25 frames per second, which amounts to roughly 135000 positions per object and match.

## 3.1 Sequence Extraction and Model Setup

In a first step, we need to identify and extract sequences corresponding to game initiations from the full game. Game initiations start with the goal keeper passing the ball. First, the goal keeper is recognized as the player with the highest average absolute position in the x-coordinate. The time of the pass is set to the moment at which the distance between the ball and the goalkeeper exceeds a predefined threshold DIST_THRESHOLD. Game initiations end with the team loosing possession, a stoppage or the start of the next game initiation as defined above. Furthermore, game initiations end after a maximum length of MAX_LENGTH. After the extraction process sequences with a length below MIN_LENGTH are excluded. In this study we set DIST_THRESHOLD = 0.1, MAX_LENGTH = 250 and MIN_LENGTH = 12, which is equivalent to 2.5-5 meters and 10 and 0.5 seconds. For clustering, the trajectories of the ball as well as the four most defensive players is used. These players usually make up the back four of the team and their behaviour during the game initiations is of particular interest to sports scientists (see [4]). The four players are associated with one group and the ball makes up a second group in the sense of Definition 1.

Figure 1 depicts two exemplary situations. The resulting numbers of game initiations for are 35 for the home team and 54 for the away team.

## 3.2 Parameter Selection

The number of clusters is choosen using the Hartigan index [5] and silhouette plots [12], which give better results than for example information criteria. This leads to 5 clusters for the game initiations of the home team and 3 clusters for the away team. For the temporal kernel we set $\sigma_T = 0.5$, which leads to some tolerance for speed differences, but still represents the ordering of the snapshots appropriately. We set $\sigma^2_{MIN}$ to the average variance of two objects' positions in a snapshot, i.e. to 0.1816, respectively to 0.2081.

## 3.3 Results

We compare the Spatio-Temporal Convolution Kernel derived in Section 2 with three baselines. First, a straight forward extension of Junejo et al. [9] to the multi-object scenario, i.e. Hausdorff distance on the set of positions of the trajectory (short: *Junejo*). Instead of the hierarchical clustering employed in [9] kernelized K-Medoids is used. The second baseline is inspired by Grimson et al. [13]. We use a Bag-of-Position as well as Bag-of-Directions representation for the the trajectories of each group. To keep it simple, we use a Multinomial Mixture Model and Expectation Maximization for clustering instead of a semantic topic model like Dual-HDP (short: *MM*). Additionally, we also compare our method with Dynamic Time Warping in combination with the product probability kernel as local distance measure (short: *DTW*).

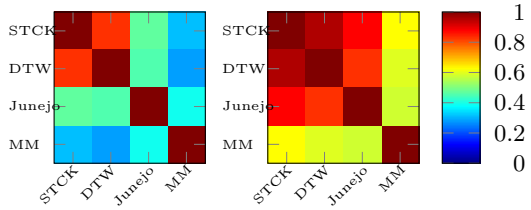


Figure 2: Adjusted rand index between our method and the baseline methods for the game initiations of the home(left) and away team (right)

Figure 2 shows the Adjusted Rand Index between the four methods. Note that in both cases the differences between the STCK and DTW are only small, while the Multinomial Mixture Model identifies significantly different clusters. Juenjo et al. finds similar clustering in one case. Figure 3 depicts the medoid, respectively the most likely trajectory, of each cluster and the average silhouette measure per cluster. The average silhouette is higher for STCK and DTW than for Junejo et al. indicating more distinctive clusters. For the the game initiations of the home team, all methods but the multinomial mixture model identify similar repre-

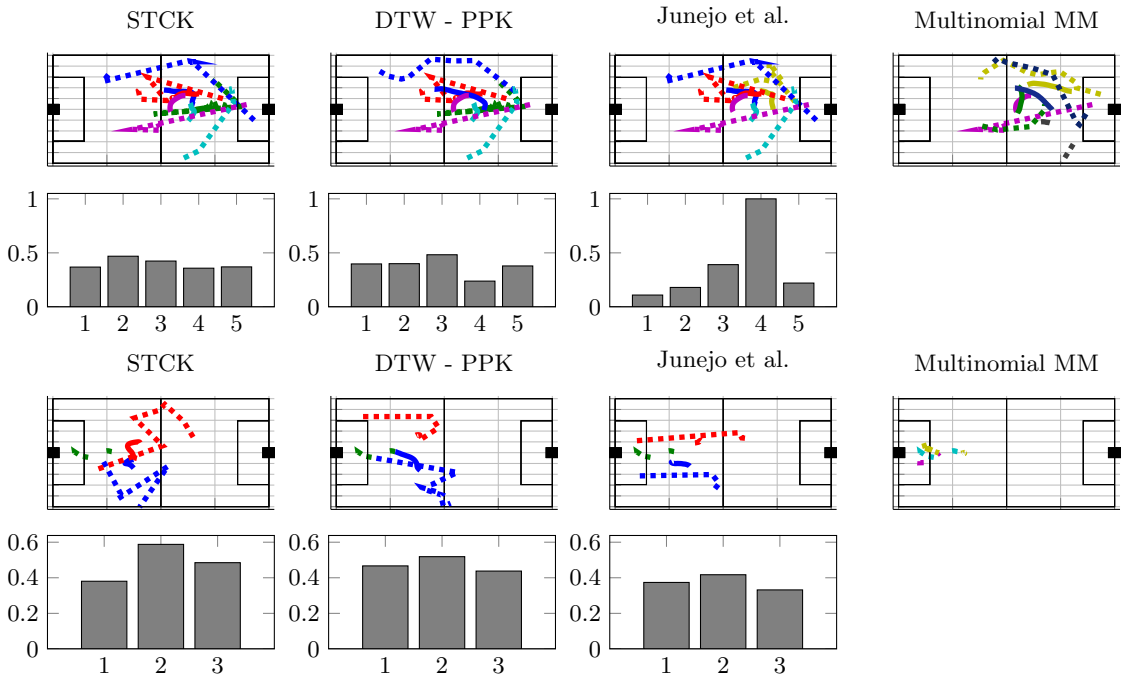


Figure 3: Medoids and average per cluster silhouette of the game initiations of the home team (top) and away team (bottom). The initiations make up 5 receptively 3 clusters indicated by the colors. The dotted line depicts the ball trajectory, the solid line the trajectory of the team centroid.

sentative trajectories. The method identifies clusters corresponding to a player carrying the ball forward on either side of the field and long (high) passes by the goalkeeper (to the left/right). Additionally, there is usually a cluster containing short game initiations, which terminate by the goalkeeper being in possession again.

## 4. CONCLUSION

We proposed novel class of $R$-convolution kernels for clustering spatio-temporal data that were composed of a temporal and a spatial kernel. The latter has been designed to represent the snapshot of a group of objects by the empirical distribution of the positions of the group members, while the temporal kernel introduced a temporal order on these snapshot distributions. Empirical results based on ball and player trajectories in soccer showed that we can achieve higher cluster separation compared to the baseline methods and identify clusters corresponding to long and short game initiations on either side of the field.

## 5. REFERENCES


[1] R. Bellman and R. Kalaba. On adaptive control processes. *IRE Transactions on Automatic Control*, 4(2), 1959.

[2] Y. Chen, A. Wiesel, Y. Eldar, and A. Hero. Shrinkage algorithms for mmse covariance estimation. *IEEE Transactions on Signal Processing*, 58(10), 2010.

[3] Z. Fu, W. Hu, and T. Tan. Similarity based vehicle trajectory clustering and anomaly detection. In *IEEE ICIP*, 2005.

[4] A. Grunz, D. Memmert, and J. Perl. Tactical pattern recognition in soccer games by means of special self-organizing maps. *Human movement science*, 31, 2012.

[5] J. A. Hartigan. *Clustering Algorithms*. John Wiley & Sons, Inc., New York, NY, USA, 99th edition, 1975.

[6] D. Haussler. Convolution kernels on discrete structures. Technical Report UCS-CRL-99-10, University of California at Santa Cruz, 1999.

[7] S. Hirano and S. Tsumoto. Grouping of soccer game records by multiscale comparison technique and rough clustering. In *Proc. of HIS*, 2005.

[8] T. Jebara, R. Kondor, and A. Howard. Probability product kernels. *JMLA*, 5, 2004.

[9] I. Junejo, O. Javed, and M. Shah. Multi feature path modeling for video surveillance. In *Proceedings of ICPR*, volume 2, 2004.

[10] L. Kaufman and P. Rousseeuw. *Clustering by Means of Medoids*. Reports of the Faculty of Mathematics and Informatics. 1987.

[11] O. Ledoit and M. Wolf. A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, 88(2), 2004.

[12] P. J. Rousseeuw. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20(0), 1987.

[13] X. Wang, K. T. Ma, G.-W. Ng, and W. Grimson. Trajectory analysis and semantic region modeling using a nonparametric bayesian model. In *Proc. of. IEEE CVPR*, 2008.

[14] X. Wei, L. Sha, P. Lucey, S. Morgan, and S. Sridharan. Large-scale analysis of formations in soccer. In *Proc. of DICTA*, 2013.