An Approach to Use Deep Learning to Automatically Recognize Team Tactics in Team Ball Games

Friedemann Schwenkreis
Dept. of Business Information Systems, Baden-Wuerttemberg Cooperative State University Stuttgart (DHBW Stuttgart), Paulinenstr. 50, 70178 Stuttgart, Germany

Keywords: Data Model, Deep Learning, Tactics Recognition, Ball Games, Video Analytics.

Abstract: Deep Learning methods are used successfully in pattern recognition areas like face or voice recognition. However, the recognition of sequences of images for automatically recognizing tactical movements in team sports is still an unsolved area. This paper introduces an approach to solve this class of problems by mapping the sequence problem onto the classical shape recognition problem in case of pictures.
Using team handball as an example, the paper first introduces the underlying data collection approach and a corresponding data model before introducing the actual mapping onto classical deep learning approaches. Team handball is just used as an example sport to illustrate the concept, which can be applied to any team ball game in which coordinated team moves are used.

1 INTRODUCTION

In case of team ball games like football or soccer, two teams of a certain number of players play against each other trying to score with a ball. In this paper we abstract from the details of scoring, i.e. we do not care whether the ball needs to be put into a goal, or a basket, or whether it needs to hit the ground in the opponent's part of the field.

One of the important features of team ball games is that they are played on a match field and the location for each active player can be determined in terms of two-dimensional coordinates relative to a specific point of the field (the origin). As an example, Figure 1 depicts the match field of team handball.

Figure 1: Match field of team handball.

Another important aspect of team ball games is the coordination of the team players – usually called *team tactics*. Sometimes individual and intuitive decisions of the players dominate the movement of players. However, the teams get trained in performing coordinated moves to improve the probability to score (offense tactics) or to decrease the probability of the opponent team to score (defense tactics).

It is an important information for coaches if and when team tactics are likely to be successful. One approach to help coaches answering this question is to manually analyze the game history based on video recordings. However, it is a very time-consuming process to analyze videos to detect team tactics and then to decide whether a certain tactical movement has been successful or not. Furthermore, for most of the team ball games it is simply impossible for humans to detect team tactics during the game due to the speed of the games.

This paper introduces a concept to detect team tactics based on sensor data during the ongoing game. Based on the concepts of location and the change thereof over time, a data representation of a team move will be introduced. Furthermore, the paper will show how to train a predictive model such that team tactics can be detected automatically, and it can be analysed whether certain team tactics are likely to be successful or not.

The concept is presented from the point of view of a data scientist. First, it will be introduced how data

Schwenkreis, F.
An Approach to Use Deep Learning to Automatically Recognize Team Tactics in Team Ball Games.
DOI: 10.5220/0006823901570162
In *Proceedings of the 7th International Conference on Data Science, Technology and Applications (DATA 2018)*, pages 157-162
ISBN: 978-989-758-318-6

can be collected in the context of team handball. Then, it will be shown, how the specific problem of automated tactics recognition can be mapped onto a well-known use of deep learning approaches. However, as being a position paper, specific results will not be presented, nor there will be a prove that the approach actually leads to a sufficient recognition rate. That will be part of subsequent publications.

2 BASICS

Basically, a team is a set of players with individual IDs (sometimes also called player numbers). Active players are differentiated from (temporarily) inactive players who are not allowed to interfere with active players or the game. The set of active players is usually called a line-up. When being part of a line-up, a player has usually an associated position or role in the line-up. However, this role of a player can change arbitrarily in some sports (e.g. this is the case for team handball).

2.1 Location and Moves of an Active Player

A team move, or a group move consists of the moves of multiple involved active players. A move of a single active player can be defined as the change of his or her position on the field over time. For team handball we have a given field geometry of 40 x 20 m which we discretize in squares of 25 x 25 cm which is a sufficient geometrical resolution for human moves in this case, because it can be excluded that two players will be in the very same square at the same time.

Based on this discretization of the field, the location of an active player can be expressed as a pair of integer coordinates, identifying the square in which the player's centre of gravity is currently located.

Since a move is a change of the location over time, we need to define a discretization of time as well. Theoretically, the maximum speed of humans can be used to calculate the maximum frequency of locations. Assuming a maximum human speed of 15 m/s and the geometric resolution of 0,25 m, we need a maximum frequency of 60 Hz to be able to resolve with the given level of detail. However, this is just the maximum frequency which will allow to detect every square that was "involved" in a player's move. If we have a lower frequency, for whatever reason, we might not be able to identify all squares a player has been in while moving from one square to another. In that case the lower frequency is depicted on the 60 Hz frequency resolution by linear interpolation between the measured locations.

2.2 Approaches to Track Players

The concept described in this paper does not depend on a specific method to detect and track the location of players on the field. However, three approaches have been investigated in the context of a proof of concept. All three approaches are suited to generate the data needed for the automated detection of team tactics.

2.2.1 Indoor-Positioning-based Approaches

Roughly, Indoor Positioning Systems (IPS) are based on the same concept as Global Positioning Systems (GPS) (Curran et al., 2011). While in case of GPS a receiver receives signals from multiple senders and calculates the differences in time the signals needed from the different senders to reach the receiver, IPS systems usually reverse that approach. A single sender sends a signal to multiple receivers and the receiver side calculates the time differences. Thus, the system can derive the position of the sender relative to the receivers. A current transmission technology for the signal exchange is ultra-wide band (UWB), as for instance used by the solutions of Kinexon (Kinexon, 2017), Catapult Sports (Catapult Sports, 2018), and other system vendors

All of the IPSs have the needed location accuracy for team handball but they differ significantly in their measuring rate ranging from 10 Hz (Catapult Sports, 2018) up to 200 Hz (von der Gruen, 2013). All the systems come with an annual cost of more than 100.000 EUR per year which is usually not affordable by most sports except for some (like soccer in Germany or football in the USA). Furthermore, the active sensors need to be attached to the players and they still have a size, which does not allow them to be used in sports where players do not wear protectors (as for instance team handball). Finally, if the position of the ball needs to be tracked as well, the ball needs to be equipped with a sender. Hence, ball vendors would need to agree on a sender technology standard for a certain type of balls.

2.2.2 Video-based Approaches

Solely video-based approaches have usually two advantages: The players do not need to wear any sensors and they are usually significantly cheaper than IPS based solutions (PlayGineering Systems, 2018). However, they have problems to keep track of the identity of players if it comes to "crowds". To reduce the likelihood of losing the identity of a player, current systems use up to eight cameras in case of indoor games like team handball, which makes these systems fairly expensive again. An alternative approach is to use a separate video system which permanently detects the identity of players by analysing the player's number, when a player enters certain areas of the field.

2.2.3 Hybrid Video-based Approaches

Combining a simple video-based tracking system (Monier et al., 2009) with a simple sensor-based identification (RFID) is a good compromise of cost and accuracy. The combined system can track the location of players in an anonymous way until the players identity is visually detected using the shirt number or when the player passes a certain RFID detection zone, which will associate an identity with the, so far anonymous, player.

In contrast to active UWB communicating sensors, RFID tags are very small and can be attached to players without violating the rules. Furthermore, they do not need any batteries and they are very cheap, in case of passive RFID tags. Unfortunately, the antennas needed to detect passive RFIDs are fairly large. Thus, the only technology that can be used in the context of team handball are antennas which are embedded in the floor or floor mats respectively.

3 DATA MODELS

3.1 Data Model for a Whole Game

We can define a data model for the location of players on a field based on an abstract definition of the data that is captured by the tracking systems. Given the introduced discretization from section 2.1, the location of a player at a given point in time is just a pair of coordinates. Consequently, the location of a team at a given point in time is the set of the locations of all active players with an associated player identification.

The location status of a whole match at a certain point in time consists of the locations of the two teams and the location of the ball, which is a set of 15 triples (player-id, x-pos, y-pos) in case of team handball and for instance 23 triples in case of soccer. To add the time dimension, we add a logical counter to the triples which indicates how many 1/60 secs have passed since the start of the game. Thus, the resulting quadruple (time, player-id, x-pos, y-pos) expresses the location of a player at a certain point in time and 216.000 quadruples are needed to encode the positions of a whole game for a single player in team handball (60 minutes playing time).

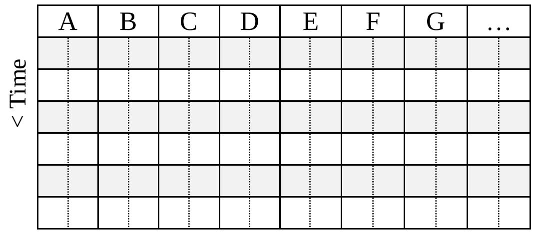

Figure 2: Player location matrix.

The quadruples can be mapped onto a 2-dimensional matrix by representing each point in time as a row with the location data of the active and inactive players and the ball. Each player identification (A to G in Figure 2) is mapped onto a pair of columns, one representing the x-coordinate and the subsequent one representing the y-coordinate of the location (see Figure 2). While the time dimension follows the sequence of locations during the play time, the player dimension has no fixed sort-order. However, we introduce some constraints (given the specifics of team ball games):

- The locations of one player must be contained in a single column which consists of pairs of values.
- If a player is inactive, then special coordinates outside the sport specific field range, (0,0) to (160, 80) in case of team handball, are assigned (e.g. 10.000, 10.000).
- The coordinates of the so-called *observed team* are mapped onto the first set of columns representing the whole team size (16 in case of team handball, including active and inactive players).
- The subsequent set of columns represent the coordinates of the opponent team (another 16 columns in case of team handball).
- The last column represents the location of the ball.

As a result, a matrix of 66 x 216.000 (60 x 60 x 60) position values represents the locations of a whole team handball match, which we denote as *match move matrix*.

3.2 Data Model for a Tactical Move

Tactical moves in team ball games are only performed by active players. Inactive players are not involved. Furthermore, we presume that substitutions are never part of a tactical move. They might happen before or after but not as part of a tactical move. Hence, a tactical move can be represented by a matrix consisting only of the locations of the active players on the field and the ball (15 columns in case of team handball).

The number of needed rows is also limited. Given the set of all possible tactical moves, the maximum duration of these tactical moves is the upper limit for the number of rows needed. The current hypothesis in case of team handball is an upper limit of 15 seconds. Thus, the maximum size of a matrix to represent a tactical move in team handball consists of 30 x 900 position values. We call this a *tactical move matrix*.

We can depict a subset of the set of rows of the match move matrix onto a tactical move matrix by omitting the inactive players. However, we need to define the criteria that allow us to decide when a player is treated as being active. The key criterion is that a player becomes part of the tactical move matrix if the player is active at the end of the subset. I.e. the location of a player in the last row in the subset is part of the range of possible locations of a match field (see section 3.1).

If we end up with less than the regular number of players in the tactical move matrix, then we will fill the empty columns with the location value of inactive players. This might be the case if one or more players have been suspended, excluded, or when they have just left the field (which can be the case in some sports). It is important to keep in mind that the order of players in a team is arbitrary with respect to the match move matrix as well as the tactical move matrix.

4 USING DEEP LEARNING TO DETECT TACTICAL MOVES

Deep Learning summarizes several pattern recognition techniques which are particularly used when we cannot explicitly find a model to describe the interrelations between certain things or actions (Goodfellow et al., 2016).

The approach described in this paper maps the detection of tactical moves onto a classification problem of video sequences. The question that we ask is: “Does a video sequence contain a tactical move or not and if so, which tactical move is it?”. It differs significantly from other applications of deep learning like face detection (for which a lot of papers have been published), because in case of tactical movements we do not focus on the similarity of single images of a whole stream but rather on the change of images over time. However, our work has a basic assumption: if we can map the sequence pattern recognition problem onto the face recognition approaches, then we can re-use the systems that have been built in that area at least to some extent.

4.1 Tactical Move Matrices and Images

Images, as they are used for face recognition, are basically just two-dimensional data structures – matrices, of colour values. Usually, the cells of the image-matrix contain three “coordinates” of a three-dimensional colour space, but there are also variants with a single value (black and white colour space) or more values (e.g. CMYB). If we think of tactical move matrices as images, then tactical move matrices are like images of a two-dimensional colour space.

There is a significant difference between images and the tactical move matrix: The tactical move matrix consists of three sets of columns (observed team, opponent team, and ball) and the position of a column has no meaning inside a column set. Thus, the columns, containing the two-dimensional values, can be exchanged arbitrarily in a column set. I.e. a matrix with a tactical move, which represents the move of player A in its first column and player B in the second column, both belonging to the same team, is semantically equivalent to a tactical move matrix that represents player B in the first column and player A in the second column (see Figure 3).

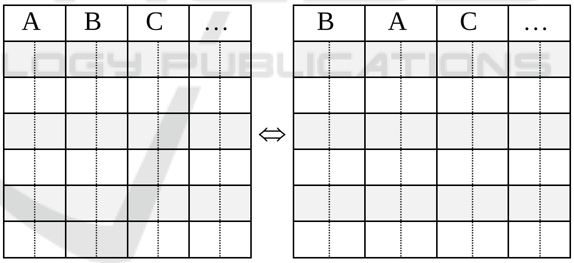

Figure 3: Tactical move matrix equivalence.

4.2 Generating Training and Test Data

With the introduced data model, the automated detection of tactical team moves can be mapped onto a classification problem. There is a limited set of tactical moves for team ball games which we define as the classes for our classification approach (in case of team handball we differentiate approximately 80 tactical moves). It is important to note that the set of tactical moves might evolve over time whenever new team tactics are invented and/or discovered. However, this is a relatively slow evolution. Thus, the model can be adapted to changes whenever there is enough data regarding a new team tactic.

Given that any tactical move matrix can be assigned to a class that denotes the tactical move which is “contained” in the tactical move matrix, training data can be derived from observed games by splitting the data of observed games into intervals which completely contain a tactical move (up to a maximum of 15 seconds in case of team handball). The splitting is done manually by searching for the end of a tactical move and then going backwards to look for the beginning of the tactical move. Then the tactical move matrix is created from the corresponding time matching rows that are extracted from the match move matrix (see section 3.2). Finally, the extracted tactical move matrix is classified with the class identifier of the previously identified tactical move.

In addition to the team move matrices, further intervals of 15 seconds are extracted from which we know that they do not contain any team tactical moves. We generate additional team move matrices from them as well to have additional test data (see section 4.4) for the “non-containing” case.

4.3 Training a Deep Learning Model

Since our objective is to use existing approaches of deep learning for face recognition, we propose to use a convolutional neural net (Goodfellow et al., 2016), CNN, to solve the classification problem. A significant difference to the approach to detect patterns in sequences of images is the fact that the associated class of a tactical move matrix is partially independent from the positions of columns in the matrix (see section 4.1).

Unless we find a canonical sort order for the columns of the tactical move matrices, we need to handle the position independence of columns explicitly. The need for “permutation invariance” of the CNN is a very similar problem to the rotation invariance in case of image recognition (Tivive and Bouzerdoum, 2006).

There are two ways to address the difference as long as CNNs do not have a built-in permutation invariance:

- Training the model with all matrices that are semantically equivalent to a given classified tactical move matrix.
- Classifying all semantically equivalent matrices when applying the model.

Since response time is critical during the later application of the model, while it is significantly less critical during training, it has been decided to use the former approach. Hence, when training a CNN with a set of tactical move matrices, the set of all semantically equivalent tactical move matrices is generated for each tactical move matrix that was generated for training. The set of the semantically equivalent tactical move matrices is derived by generating all permutations of column positions for each player inside a team. I.e. in case of team handball, we have 7! = 5.040 permutations for each team and $(7!)^2$ matrices which are semantically equivalent.

4.4 Testing and Applying the Model

The resulting model after the training phase is tested using additional pre-classified data containing a tactical situation, as well as tactical move matrices which do not contain a tactical move. The classification result for every tested tactical move matrix is a single “class association” which is generated by the final activity function of the CNN. This predicted class is then compared to the previously assigned class value. We use a classic “domain specific confusion matrix” to measure the quality of our model, which means that we weigh errors according to their severity in the domain (team handball in our case).

The application scenario of the model is based on the constant stream of location information during a match or even a training session. We are periodically extracting tactical move matrices from the stream containing the location records for the maximum length of a tactical move (the assumed 15 seconds for team handball). These tactical move matrices are then sent to the model for classification to detect whether a team tactical move was performed.

5 CONCLUSION AND OUTLOOK

This paper describes the work of an ongoing project. Using the example of team handball, we introduced an approach to automatically detect team tactics using a deep-learning-based classification model. In particular, we have defined the necessary data models and transformations. Given the available sensor technology for location detection of players, the described approach can be used to train a convolutional neural network.

Although the concept has been defined, there are still a number of open points to be worked on:

- Developing a permutation invariant CNN to avoid the generation of semantically equivalent matrices to train a model.
- Finding the appropriate activation function to generate the predicted class.

- Assessing the severity of different categories of misclassifications
- Finding the optimal frequency for the extraction of team move matrices for classification. Currently, we assume that a classification frequency of 1 Hz will be sufficient.

With the “online” availability of the player’s location data, being detected with the necessary accuracy, a complete new data science view of team ball games becomes possible. The performance of players can be described in a completely new way, which takes the movement of players and their relative position into account.

Furthermore, the automated detection of team tactics is the basis for the automated prediction of the success of team tactics as well as for the definition of a new class of player performance indicators. The world of coaches will further change if they can base their decisions on information regarding whole teams rather than on player specific values only.

REFERENCES

Catapult Sports. (2018). *ClearSky T6*. Retrieved from https://www.catapultsports.com/products/clearsky-t6

Curran, K., Furey, E., Lunney, T., Santos, J., Woods, D., & McCaughey, A. (2011). An evaluation of indoor location determination technologies. *Journal of Location Based Services*.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.

Kinexon. (2017). *Real-time Performance Analytics.* Kinexon.

Monier, E., Wilhelm, P., & Rückert, U. (2009). A Computer Vision Based Tracking System for Indoor Team Sports. *The fourth International Conference on Intelligent Computing and Information Systems.* Cairo.

PlayGineering Systems. (2018). *Pioneers in sports` technologies*. Retrieved from http://playgineering.com

Tivive, F., & Bouzerdoum, A. (2006). Rotation Invariant Face Detection Using Convolutional Neural Networks. In I. King, J. Wang, L. Chan, & D. Wang, *Neural Information Processing. ICONIP 2006* (S. 260-269). Springer.

von der Gruen, T. (2013). Retrieved from Funkbasierte Lokalisierungstechnologien – RedFir & Co – Ortung im Sport, am Flughafen und in der Logistik: http://www.angewandte-kartographie.de/download/symposium2013/vortraege/Von_der_Gruen_RedFIR.pdf