Tactics Analysis in Soccer – An Advanced Approach

Jürgen Perl², Andreas Grunz¹ & Daniel Memmert¹
¹German Sport University, Cologne, Germany
²Johannes Gutenberg-University, Mainz, Germany

Abstract
In order to run a game tactically, high level knowledge is required from and by coaches and analysis experts. Assessing tactical performance through statistic indicators has some drawbacks, however, it is not only difficult to prove reliability on the defined indicators, but those static indicators often hide the game's dynamics. New network-based approaches offer a promising way to improve future evaluation of tactical performance and recognition of game dynamics. The aim of this article is to introduce a new approach where pattern-based tactics analysis is combined with success-oriented statistical frequency analysis. Therefore, the neural network-based pattern analysis of the SOCCER approach (Perl & Memmert, 2011) has been completed by an event-oriented statistical analysis, which is mainly based on an automatic recognition of ball possession as an indicator of success. After a short introduction into basic aspects of game analysis, automatic position tracking and net-based pattern analysis, the new concept of SOCCER is presented in two steps: The first part deals with net-based analysis of dynamic processes, oriented in constellations of tactical groups of players. The second part deals with rule based semantics analysis, which allows automatic recognition and evaluation of individual activities and embedding them into the tactical patterns - thus enabling both, evaluation of tactical processes based on individual success as well as evaluation of individual activities in the context of tactical processes.

KEYWORDS: COMPUTER SCIENCE, GAME ANALYSIS, SOCCER, PATTERN RECOGNITION, NEURAL NETWORKS

Introduction
In the quarterfinal of the FIFA World Cup 2006, Germany versus Argentina, the goalkeeper of the German national squad, Jens Lehmann, paved the way into the semi-final. Information about shooting preferences of his opponents, gained by the analysis of their penalty striking habits, helped Lehmann in the decisive shootout to choose the right corner and to perform saves (Buschmann & Nopp, 2006). At the FIFA World Cup 2010 it was determined that the opponents of the English squad operated with an increasing number of long passes, which revealed deficits in the English defense (Buschmann & Nopp, 2010). In the round of sixteen in the match against England, the German squad took advantage of these findings. The goalkeeper Manuel Neuer made an assist by passing a goal-kick over the English defense line to forward Miroslav Klose who scored the lead. The examples highlight how performance analysis in soccer can be a key factor in effective game preparation, in addition to the
monitoring and evaluation of training (for a review, Nopp, 2012).

On the one hand, spectacular cases like those are impressive. They simplify how important it is to get key information out of the game. On the other hand, they demonstrate the large gap between isolated key information and the knowledge on complex tactical behaviour. In order to recognize tactical plans and their success, it is necessary, based on recordable data, to recognize and analyse the behavioural patterns of the players and in particular of the tactical groups. Only if it is understood, in which interaction what constellation or move of a tactical group is successful and opens ways for individual tactical maneuvers, the tactical orientation of the team and the individual players can be improved.

Since the early 1970s, computerized notational analysis has been playing an important part as a scientific basis for developing concepts of recording and analyzing data from games. There are some scientific studies that improved the understanding of soccer by underpinning soccer performances analysis with theoretical findings (Memmert & Perl, 2009b, for a review). Hughes & Franks (2005), for example, clarified the need for normalization of data – the latter research group counted the number of passes per possession and linked that to resulting goals; due to unequal occurrences of each possession length (according to passes) a normalization of the data was inevitable. Therefore, Hughes & Franks (2005) divided “the number of goals scored in each team possession by the frequency of the sequence length” (p. 511), multiplied the results by 1000 in order to avoid small ratios – and by doing so, highlighted differences between successful and unsuccessful teams considering the style of play (possession play or direct play) and resulting conversion ratios to shots on the goal. This was based on another far-reaching study conducted in 1968 by Reep & Benjamin. This study coined the British and Norwegian soccer team’s style of play by arguing that direct play – implying few passes and a high frequency of shots on the goal; thus playing highly penetrative – results in successful outcomes.

Obviously, such data is important to receive information about a lot of performance indicators reaching from technical skills over physical condition to most relevant skills like the ability of scoring goals.

However, most of those distributional and frequency oriented results, although doubtlessly helpful in practice, are primarily useful for classification and ranking, but neglect the dynamic aspects of processes like interaction and context. They "freeze" the ninety-minutes-game to just some handful of numbers, helping to understand "what" but not "why". The key for a better understanding of the game is to analyze what the coach is doing: He "reads" the game, i.e. he does not record numbers but recognizes patterns.

Therefore, the aim of this article is to give an introduction into methods and first results of pattern-based tactic analysis, which has been successfully run by means of artificial neural networks approximately during the last 4 years. Although it is not perfect at all, it shows that those networks can help to map the complex game to a sequence of patterns. Moreover, those patterns can be combined in a fruitful way with statistical results and/or the patterns themselves can be used for statistical analyses.

In the following, the emphasis is put on three main aspects:

The best fitting data for description and analysis of spatio-temporal processes like those in soccer is given by the positions of the players and the ball. Therefore, the first part deals with automatic position data recording. A brief overview introduces the most interesting up-to-date approaches.
In order to analyse the game processes in a qualitative way, the focus is laid on patterns of actions and interactions. Some main aspects of pattern analysis, in particular by means of artificial neural networks, are briefly introduced in the second part.

The third part introduces a net-based software-tool "SOCCER", which combines net-based pattern analysis with rule-based semantics, statistical analyses and expert knowledge to gain greater insights into the game dynamics. Moreover, it is very important to make data easily understand considering the data output. SOCCER, like most of the analysis software systems available, has integrated graphics, which help in terms of data interpretation. Whether the performance is coded in-event or post-event, the data is reinterpreted based on edited videotapes or graphical illustrations calculated by the software. Modern systems make it possible to refine actions in different areas of the pitch or at certain intervals of the game, e.g. during a certain match status. Finally, data can be used to compile statistics that can help to identify performance profiles of players and/or teams as well as the strengths and weaknesses.

Position Tracking in Soccer

In computerized notational analysis, distances covered during a game are of interest to quantify the intensity of motion. The amount of walked, jogged and sprinted meters as well as the velocity and acceleration of each player during a game are indicators for the condition of the team. In order to analyze actions and interactions, first of all the constellations of the players and the types of moves are of interest in computer-based tactics analysis. At any case, the positions of players and the ball play the important role and therefore have to be recorded from the game.

Given that video tracking is available in the majority of competitive games, research has focused on methods to extract position data of the players from recorded videos. In 1990, Herzog & Retz-Schmidt proposed a tracking system using image processing. They used a fixed camera that covers the entire pitch to record image data. This is a software programme that simply instructs the computer to take the current position, in (x, y)-pixels on the screen at any frame, and to calculate the distance travelled since the (x, y)-pixels at the previous frame. A first commercial prototype, the computer vision system AMISCO, was released in 1998 by Videosports Ltd. to obtain position data. Since then, the method of image processing made a big jump ahead, however, it still needs a supervisor to control and correct data. Impellizzeri, Sassi & Rampinini (2006) showed that the variation of different supervisors is below 2% in one tracked game.

More recent systems are based on frequency modulated continuous wave (FMCW) technology to track positions. The most common representative of this technology is the GPS system. Players have to carry a transmitting unit while playing, to be detected by the system. A comparison of accuracy of GPS and video based systems showed an overestimation of 4.8% for GPS and 5.8% for video examining distances during games (Edgecomb & Nortona, 2006). At present, the LPM system produced by ABATEC has the lowest distinction of just 1.6% underestimation of real and measured distance (Stelzer, 2004). It uses FMCW sensors that send a permanent signal to a certain number of receivers around the pitch. A central processing unit collects these data and is able to present it in real time. Although FMCW systems a more accurate and faster in presenting the data, they can just be used in training due to FIFA regulations at this time.

The position data collected with the different approaches in particular can be used to fulfill net-based tactics analysis, presented in the following.
Pattern Recognition

Pattern recognition based on position data can result in different kinds of patterns (Memmert & Perl, 2009a):

The formations of (tactical) groups of players – i.e. the positions of the group members in relation to each other – build spatial patterns. The time-dependent movements of such groups build temporal patterns. Both types of patterns help to recognize tactical concepts. In combination, these patterns build spatio-temporal patterns of the game processes. And the combination of respective patterns of both teams result in interaction patterns, which are helpful in order to measure the success of tactical actions in the context of tactical interaction.

The patterns, if once obtained, can be used to calculate several statistics like frequency or rareness of action or interaction patterns. In combination with additional semantic information like success/failure, this can already give deeper insights into the game dynamics. In our research we could identify several rare patterns which also had a high probability of success. The terms "success" and "failure" can be defined in several ways depending on the research question (compare the section "Rule based semantics analysis").

To obtain such patterns, a lot of methods are available, stretching from simple similarity analysis over statistical clustering methods to neural network approaches. Basing on a lot of positive experiences with sports as well as with technical applications, we decided to use neural networks – in particular because in the case of self-organizing maps (see explanations below), no pre-information about number and types of clusters is necessary.

The applied methods can be divided in supervised and unsupervised methods. The term supervised/unsupervised has its origin in the scientific field of machine learning. Supervised methods learn patterns by examples. For instance the group tactic \textit{wing play} in soccer can be learned by feeding the net with examples of \textit{wing plays}. In contrast, unsupervised methods are characterized by the lack of labeled data. A similarity measure is used to group the data and to construct distinct prototypes. While in supervised learning a predetermined pattern is learned by examples, in unsupervised learning there are no predefined patterns. Therefore, the patterns gained by unsupervised methods do not necessarily correspond to conventional standard patterns like \textit{wing play}, but surprisingly often do: If a pattern is sufficiently represented by the data and distinct from other ones, the net will normally recognize it by its own.

The self-organizing map (SOM) developed by Kohonen (2000) is an artificial neural network of the unsupervised type. A SOM consists of a set of artificial neurons that are connected to each other through edges usually arranged in a rectangular grid. During the training phase the network adapts itself to the distribution of the data used for training. After training each neuron encodes a different pattern. In an additional step, the neurons encoding similar patterns are grouped into clusters, representing a pattern-prototype.

The kind of patterns represented by the network is determined through the data that is used for training. If the network is trained with movements of one group of players, the resulting patterns will encode typical movements of that group. If the network is trained with interactions, consisting of movement data from 2 interacting groups and the ball, the resulting patterns will encode typical interactions.

The Dynamical Controlled Network (DyCoN; Perl, 2004), which is derived from SOM and overcomes several technical limitations of the original SOM-concept, has successfully been used to detect tactical patterns in soccer games (Memmert & Perl, 2009a, 2009b; Grunz, Memmert & Perl, 2012).
An example for supervised learning of specific tactical patterns in soccer like game initiations is given in Grunz et al. (2012). A hierarchy of neural networks was developed to learn these patterns. As the hierarchy is trained with labeled data gained by an expert watching and categorizing the game, the hierarchy belongs to the supervised methods. Several sets of example data extracted from the categorization were used to learn the corresponding patterns. For instance, the expert categorized a set of sequences that showed short game initiations. The position data corresponding to these sequences was extracted by a program to train the hierarchy of networks. After training the hierarchy, it was used to classify tactical patterns in new games. While an expert needs at least 5 hours to manually categorize a game, the developed approach can reduce the required time to a few minutes.

**Combination of Pattern Recognition with Rule-Based Semantics and Statistical Analysis**

This net-based game analysis approach in soccer is mainly based on the ideas of data reduction (Perl, 2008), pattern recognition (Grunz, Memmert & Perl, 2009; Memmert & Perl, 2009a, 2009b; Grunz, Endler, Memmert & Perl, 2011), the network DyCoN (Perl, 2004; McGarry & Perl, 2004), and the special game analysis software SOCCER (Perl & Memmert, 2011), which all work together as it is briefly described below.

The process starts with position data preparation and pre-processing, which is done by means of the software tool SOCCER, followed by three steps of analysis:

1. The position data of the players of a team are reduced to those of tactical groups like offense or defense, followed by normalization to standard patterns, as it is shown in Figure 1.

2. The net is trained with those formations, resulting in a collection of formation clusters, each containing a collection of variants of the corresponding formation type.

3a) Along the time-axis, position data of interacting tactical groups are fed to the net, which recognizes the time-dependent corresponding formation types as well as in particular striking features.

3b) Additionally quantitative analysis of frequency distributions of formation types is done by means of the statistics tool of SOCCER.

3c) The trajectory analysis component of DyCoN enables tactical analyses of the game, including interaction and phase analyses. In particular long term interaction patterns as well as hidden or creative tactical activities can be recognized and analyzed regarding success.

![Figure 1. Departing a constellation into its position (centroid) and its characteristic formation (taken from Perl & Memmert, 2011).](image)
Those time-series of formation types together with positions and statistical information allow a wide range of analyses from space- or time-oriented distributions over success in interactions up to tactical aspects. Not least, the condensed game information from above can be used for generating a game protocol, where time, formation type and position data can be completed by semantic information. Such a protocol then builds the basis for the actual game analysis, as it will be discussed in the following.

**SOCCER-Based Game Analysis**

The basic concept of SOCCER is to handle two types of data: On the one hand syntactic and semantic items taken from video and automatic position recording as well as from expert evaluation, and on the other hand patterns of formations and formation sequences taken from net-based analysis. Recorded data as well as recognized patterns can be analyzed statistically under the aspects of frequencies and spatio-temporal distributions. The central idea of the approach is that both groups of information can be combined in a compound analysis: The formations and formation sequences build the basis for the understanding of interaction and tactical patterns. They are also useful as a background and/or context for the evaluation of statistical items. In turn, syntactic and semantic items are helpful for understanding and evaluating pattern constructs.

Original data, patterns and results from analyses are organized in a data base. They can be presented in interactive tables, graphics and animations. This allows an arbitrary combination as well as a stepwise resolution of the presented information.

---

![Diagram](image)

Figure 2. Basic concept of the SOCCER analysis tool (list of items is only a selection).

In the following, three examples of SOCCER-based analyses are presented, followed by a closing section on rule-based semantics analysis:
Distribution analyses. The distribution analysis presented in Figure 3 demonstrates a typical situation: Team A attacks in the formation of type 4, team B reacts with a defense formation of type 3. The distribution matrix shows that this particular interaction happened 523 times (i.e. at 523 seconds) in the corresponding half-time.

In general, the matrix provides the distributions of formations of the teams as well as those of the respective interactions.

Statistical analysis is helpful for a first recognition of normal and of seldom or striking situations. In order to recognize the role they play in the game process, statistical analysis can be combined with animated process analysis.

Combined quantitative and qualitative process analysis. As mentioned above, the formation data can be completed by semantic ones like technical or tactical aspects and success. The following example deals with evaluating the success of a team in a given formation interaction. Figure 4 shows from left to right the number of evaluated interactions of a team, followed by the negative ones in absolute numbers and as percentages. Concentrating on the right graphic, it seems that team A has serious problems in the interaction of formation 3 vs. formation 3. However, the absolute numbers are very small, reducing the importance significantly. Also ’5 vs. 5’ is negative but does not seem very important, whereas ’5 vs. 2’ seems to be a significant weakness, although the percentage of negative results is only 16. Note, that the presented analyses are only examples, which can be completed arbitrarily if the valuation data is once available.

Net-based tactic analysis. Tactic analysis is done by net-based trajectory analysis. The idea is that at each point in time the formations of tactical groups are identified and can therefore be encoded by a corresponding number and/or color. After training, the network can recognize the
formation and the formation type contained in each data set of the original position data, and is therefore able to map the original process to a trajectory, as it is demonstrated in Figure 5.

The graphic shows a net of neurons, i.e. the small white or colored squares, where each color stands for a formation type like the one in Figure 1. Different neurons of the same color represent variant formations of the same type. Representing those variants by just one characteristic type reduces the number of significantly different items to only about 10, which has two important advantages: On the one hand, it enables statistical analyses on reasonable distributions. On the other hand, the formation trajectories are smoothed and therefore allow easier comparisons between each other. (Note that all specific information is saved and can easily be used for special analyses if needed.)

Figure 5. A trajectory of formations on the net and its reduction to a formation type trajectory.

In the presented example from Figure 5 it works as follows: The position data sets of the game process activate corresponding neurons of the network, starting with the one with the red mark. The process then runs through some light green neurons followed by some orange and some red ones and so on. Reduced to the significant types represented by the corresponding colors, the trajectories become much simpler and therefore represent the specific behavior of the corresponding tactical group (see the small embedded graphic on top left).

Such tactical phase patterns can be put in, clustered and finally recognized using neural networks on a second level. In the following, an example of a striking feature analysis is given which demonstrates the way, how a complex analysis of combining quantitative and qualitative aspects works: Figure 6 presents a sequence of corresponding formations of team A and team B. A first glance on the formation phases shows that team B has frequent changes from 2 to 4 and back, while A has analogous changes between 5 and 3. These correspondences can be systematic or arbitrary. A second level analysis may help to answers questions like this and lead to a better understanding of such tactical interactions.
Rule-Based Semantics Analysis

A last and most challenging task is the one of automatically recognizing semantic information from the game data: Although formations and tactical interactions play an important role for the understanding of the processes in the large scale, the practice of players and coaches demands information on the ball win and loss, ball possessions, starts of attacks and success of offense or defense activities in general. In turn, that particular information on actions improves its value and importance by far, if embedded in the context of the behavior of the involved tactical groups.

The problem, however, is how to get the semantic information from the position information only. The basic idea of one of the authors (Perl) is that the position data of the players and the ball give information about the probability that the player closest to the ball is in contact with it, if the distance between that player and the ball is smaller than a sufficiently small distance. Of course, this assumption is not correct every time, but from the statistical point of view, it helps to deduce the following information: (a) Ball win / loss: The ball contact changes from a player of one team to the other team; (b) Ball possession: Over a certain time interval, players of the same team have ball contact; and (c) Making / receiving a pass: The ball contact changes from a player of a team to a (normally) distant one of the same team. To evaluate single activities, "win", "making", "receiving" and "possessing" can be taken for "successful", while "loss" or "without ball contact" can be taken for "unsuccessful".

Based on this elementary information, processes and their evaluations can be defined. Starting and running an attack, for instance, can start with the ball win, followed by a ball possession, and followed by a pass and so on, normally ending with a ball loss. The final loss, however, naturally does not mean that the whole process was unsuccessful, i.e. only the steps are counted. Finally, positions and contact measuring help to evaluate the tactical behavior of groups or the whole team: One example is the compactness of a team relatively to the ball, which gives important information on the players' positioning and ball orientation.

A second example is the speed of a defense process in getting the ball under control and leading it back to start an attack. A third example is the speed of an offensive process in getting the ball into the opponents half. The SOCCER data preparation and management offers a complex handling of all information about players, their positions, formations, ball contacts and success. This is organized by means of time depending on data vectors which are – based on the integrated data base concept – reflected in interactive tables on the interface. Figure 7 shows how it works.

Every combination of players, groups, formations, ball contact situations, areas on the soccer pitch, and success can be activated just by clicking the regarding input tables, resulting in
information about when, where, how successful and how often (absolute, in percentages, compared to the opponent team) the selected situation was. In the example in Figure 7, the analysis focused on the ball possession activities of the offense of team A in the area close to the goal on the right hand side. The result was that formation 3 was 50% dominant in the corresponding points in time t and the formation sequences over the respective last 6 seconds were rather constant (see table on right hand side).

Figure 7. Example of a database inquiry and result interpretation.

Figure 8 demonstrates a specific analysis dealing with ball losses in the context of formation and areas. The result of the presented example is that team A has a symmetric right-left distribution of losses over all formations with a minimum of only 2 losses in the context of formation 4. It should be mentioned, however, that the absolute number of formation 4 is also rather small, i.e. the percentages, which can also be retrieved from the database are an important measure for the ball possession value of formations.

Finally, speed and compactness analyses are special features of SOCCER: The subject to the analysis is the process where the defence stops the opponent attack and passes the ball as far as possible to the own offence in the opponent's half. The process stretches over four points in time, t0, t1, t2, t3, beginning with the ball win and ending with the offence's attack, and is analysed regarding the contexts of formation, tactical groups and specifically involved players. This way, two main questions can be answered:

(1) What is the speed of the process along t0, t1, t2, t3 – i.e. how fast can the team or tactical group react and start counter-attacks?
(2) How does the centroid of the team or tactical group moves related to the ball, and how does its compactness change?

Experiences from the analysis of more than 40 games shows that only the combination of position-based analysis and semantic analysis in the context of tactical group formations is able to make complex playing processes transparent and understandable, helping to improve tactical behaviour.

**Conclusion and Outlook**

Over the past years, progress of computer science made it possible to track the players’ movements and thus to provide position data. Neural network approaches have become a frequently studied and commonly recognized possibility for data analysis and data simulation in sports (Memmert & Perl, 2009b). These studies have demonstrated that additional research questions linked with pattern learning, game analysis and simulation processes can profit by using neural networks. All in all, the role and development of neural networks is now a topic of current discussions in computer sport science and also in performance analysis in soccer.

In the field of game analysis, SOCCER became a tool for automatic analyses and assessments of tactical behavior based on position data for the first time. If validity can be ensured, the planned assessment system will be an important step towards objectification of tactical performance components in team sports. Furthermore, there will be a dramatic speed advantage concerning the evaluation of the position data (from 6-8 hours to 2 minutes). The small effort for data acquisition will enable the accumulation of a vast amount of data and will thus bring new chances for theory construction and practice in soccer.

**References**


