T-pattern analysis in soccer games:
Relationship between time and attack actions

Detección de T-pattern en los partidos de fútbol:
Relación entre las acciones del equipo y de ataque

Análise T-pattern em jogos de futebol:
relação entre as ações do tempo e de ataque

Cesare Cavalera¹, Barbara Diana¹, Massimiliano Elia¹, Gudberg K. Jonsson², Valentino Zurloni¹ y M. Teresa Anguera³

¹University of Milano-Bicocca (Italy), ²University of Iceland (Iceland), ³University of Barcelona (Spain)

Abstract: Sports performance consists of a multiple series of strategies that tend to follow one another. Performance analysis in team sports is usually focused on primary (fundamental skill execution), secondary (scoring) and tertiary (match result) outcomes. While there is general agreement over measuring secondary and tertiary outcomes, literature does not show a unanimous agreement over a unique measure of the primary level of performance. The aim of this study was to investigate primary performance outcomes through an analysis of temporal patterns. In particular, we were interested in verifying if changes in tertiary performance outcomes may be related to changes in primary ones. We selected three soccer matches played by a top club during the Serie A league over the 2012-2013 season in which there was a change in match result between first and second half (tertiary level of performance). The methodological approach was based on observational design, supported by digital recordings and computer analysis. Data were analyzed with Theme 6 beta software, which detects the temporal and sequential structure of datasets, revealing repeated patterns that may regularly or irregularly occur within a period of observation (T-patterns). Striking differences were found comparing first and second half temporal patterns, especially when the final match outcome showed an improvement of the first half’s one. Our results suggest that Theme software and T-pattern enhance research opportunities by identifying a useful tool to study the link between primary and tertiary level of performance, making this an effective research and support instrument for sports analysis.

Keywords: Soccer performance; observational methodology; T-pattern detection.

Resumen: El rendimiento en el deporte viene condicionado por múltiples estrategias que tienden a sucederse entre sí. El análisis del rendimiento en los deportes de equipo se centra especialmente en resultados primarios (ejecución de competencias básicas), secundarios (marcadores al puntuar) y terciarios (resultados de los partidos). Mientras que existe un acuerdo general acerca de cómo medir los resultados secundarios y terciarios, la literatura no es tan unánime respecto a hallar una única medida relativa al nivel primario de rendimiento. El objetivo de este estudio es la investigación acerca de estos resultados primarios de rendimiento mediante un análisis de patrones temporales. En concreto, estamos interesados en verificar si los cambios en resultados terciarios están relacionados con los primarios. Seleccionamos tres partidos de fútbol que se jugaron en un club situado en muy buena posición durante la liga de la Serie A en la temporada 2012-2013, en la cual se produjeron cambios en el resultado entre la primera y segunda mitad de los partidos (nivel secundario de rendimiento). El planeamiento metodológico se ha basado en un diseño observacional, que se ha desplegado mediante el apoyo de registros digitales y análisis informatizados. Los datos se analizaron mediante el programa informático Theme 6 beta, que permite detectar la estructura temporal y secuencial de los datos, detectándose patrones repetidos que pueden ocurrir de forma regular o irregular en el período de observación (T-patterns). Se hallaron diferencias sorprendentes al comparar los patrones temporales de la primera y segunda mitad, especialmente cuando el resultado final del partido mostraba una mejora del marcador respecto a la primera mitad. Nuestros resultados sugieren que el programa Theme aporta posibilidades de mejora en la investigación, convirtiéndose en un instrumento útil para estudiar la conexión existente entre los niveles primario y terciario de rendimiento, a la vez que se demuestra que es un importante apoyo para una investigación eficaz en el deporte.

Palabras clave: Rendimiento en fútbol, metodología observacional, detección de T-patterns.

Resumo: O rendimento no desporto está condicionado por múltiplas estratégias que tendem a suceder-se entre si. A análise do rendimento nos desportos de equipa centra-se especialmente nos resultados primários (execução de competências básicas), secundários (pontuação) e terciários (resultados dos jogos). Embora exista um consenso geral acerca de como medir os resultados secundários e terciários, a literatura não é tão unânime no que concerne a uma única medida relativa ao nível primário de rendimento. O objectivo deste estudo é a investigação acerca destes resultados primários de rendimento através da análise de padrões temporais. Em concreto, estamos interessados em verificar se as mudanças nos resultados terciários estão relacionadas com os primários. Seleccionámos três jogos de futebol de um clube em muito boa posição classificativa durante a liga Serie A na temporada de 2012-2013, no qual se produziram alterações entre os resultados entre a primeira e a segunda metade dos jogos (nível secundário de rendimento). O planeamento metodológico baseou-se num delineamento observacional, que se levou a cabo com o apoio de registos digitais e de análises informatizadas. Os dados foram analisados através do programa informático Theme 6 beta, que permite detectar a estrutura temporal e sequencial dos dados, detectando-se padrões repetidos que podem ocorrer de forma regular ou irregular no período de observação (T-patterns). Foram verificadas diferenças surpreendentes ao comparar os padrões temporais da primeira e segunda parte, e especialmente quando o resultado final do jogo revelava uma melhoria do marcador relativamente à primeira parte.
Introduction

Performance outcomes are widely acknowledged sporting variables, but have received very little scientific attention. Usually, three factors have been distinguished (Courneya & Carron, 1992): a primary one, corresponding to the fundamental level of performance (such as penalty kick goals in football or free throw percentages in basketball); the secondary level corresponds to the intermediate or scoring aspect of performance (such as goals scored in football or points scored in basketball); the tertiary factor is the traditional outcome measure related to win or loss ratio.

Most football studies have investigated the link between performance and variables such as game location (home/away; crowd, travel) (Brown Jr et al., 2002; Courneya & Carron, 1992; Tucker, Mellalieu, James, & Taylor, 2005; Page & Page, 2007; Lewis & Goltzi, 2007; Pollard & Seckin, 2007; Gómez, Lago, & Pollard, 2013) and critical psychological and behavioral states (competitors, coaches, officials) (Neave & Wolfson, 2003; Wolfson & Neave, 2004; Pollard, 2008). Only few studies (Yiannakos & Armatas, 2006; Armatas, Yiannakos, & Sileloglou, 2007; Armatas, Yiannakos, Galazoulas, & Hatzimanouli, 2007; Armatas, Yiannakos, Papadopoulou, & Skoulas, 2009; Jinshan, Xiaoke, Yamanaka, & Matsumoto, 1993; Michailidis, Michailidis, Papaiaakovou, & Papaiaakovou, 2004) have investigated the link between time and performance in football comparing performance outcomes between first and second half of the matches.

Most of these studies showed that the frequency of primary outcomes such as goals scored is time dependent. Analyzing Euro 2004 matches, Yannakos and Armatas (2006) showed that 57.4% of goals were scored in the second half (p<0.05), underlying a systematic and significant upward trend of goals scored as time progressed. Similarly, a study that analyzed correlations between time and goal scoring in the three latest World Cups (Armatas, Yiannakos, & Sileloglou, 2007) showed that more goals were scored in the second half of the games and, in particular, in the last 15-minute periods. These data are consistent with another study (Armatas, Yiannakos, Galazoulas, et al., 2007) that examined the same correlation in the three latest women’s World Cup. Furthermore, a study that analyzed 240 matches from the Greek “SuperLeague” 2006-2007 (Armatas et al., 2009) revealed that more goals were scored in the second half of the match with a systematic upward trend in the number of scores as the match progressed. Apart from the number of goals scored, other important factors of primary outcome were investigated by Rampinini and colleagues (2009), who compared technical performance of the first five teams with the last five teams of 2004-2005 Italian Serie A. Results showed a significant decline (p<0.01) between the first and second half related to involvements with the ball, number of short passes, and successful short passes in both more and less successful teams. These results are mostly attributed to physiological factors of the game. Consistent with this point of view, there is a great amount of knowledge supporting a reduction in physical condition over the course of the match (Bangsbo, 1994; Krustrop, Mohr, Ellingsgaard, & Bangsbo, 2005; Krustrop et al., 2006; Mohr, Krustrop, & Bangsbo, 2003; Rampinini et al., 2007; Saltin, 1972). These studies showed a greater state of fatigue and reduced physical performance during the course of the match. In particular some studies examined physical activity of both English Premier League soccer players (Rampinini, Coutts, Castagna, Sassi, & Impellizzeri, 2007) and referees (Weston, Bird, Helsen, Nevill, & Castagna, 2006) showing a decreased physical performance in the second half. Consistent with the results offered by Rampinini and colleagues (2009), these findings suggest a link between professional physical fitness and the fatigue-induce worsening of technical and tactical skills. Furthermore, a study by Helgerud and colleagues (2001) showed an increase in the involvements of the ball and a decrease in unsuccessful passes after 8 weeks of aerobic interval training. These data can be explained considering that players display a greater deterioration in terms of physical condition and lapses in concentration that can be related with differences in performance outcomes (Reilly, 2003).

Besides this great amount of data that suggest a link between time and football performances, other studies support the evidence that there is no immediate correlation (Jinshan et al., 1993; Michailidis et al., 2004). Jinshan and colleagues (1993) in particular, found three periods of play (0-10, 30-40 and 70-80 min) producing lower amounts of shots than other times. The reason for this could be that these times of the match are considered to be bad times to concede goals, which could cause teams not to commit as many players forward in fear of a counter-attack, thus restricting attacking options. Nevertheless, such an explanation doesn’t take into account other features of the game, such as the momentary score or the number of goal difference that may contribute in explaining the absence or presence of offensive patterns.

From these data it is crucial to provide a more thorough reading of the offensive strategies related to the particular moment of the match. In line with these needs, the most
recent studies in the field of sport have underlined the importance of identifying visual undetectable structural regularities in order to better assess the complex reality they refer to (Anguera, Camerino, Castañer, & Sanchez-Algarra, 2014; Camerino, Chaverri, Anguera, & Jonsson, 2012; Jonsson et al., 2010; Zurloni, Cavaleria, Diana, Elia, & Jonsson, 2014). One approach that can meet these needs involves the detection of temporal patterns, known as T-patterns, in order to reveal hidden yet stable structures which underlie the interactive situations during football matches. Studies based on the T-pattern method have been conducted in very different sport domains apart from football, such as boxing, basketball, swimming and handball (Jonsson et al., 2010). This methodology is based on observational design, allowing the analysis of all the relevant information about any aspect related to any interactions linked with primary factors. Due to recent debates related to the link between performance outcome and time, more studies are needed in order to better explore differences related to tactical and technical strategies during different times of the match. Therefore, this paper aims to detect hidden patterns related to offensive game comparing performance outcomes of first and second halves of football matches. Since there is a greater degree of predictability when referring to the same team, our first aim is to reveal the hidden yet stable structures which underlie the dynamic situations (Shepherd, Lee, & Kerr, 2006) that determine the attack actions of play in different soccer matches of the same team. Secondly, to understand whether and how the temporal and sequential organization of attack actions can affect the outcome of the performance, we compared the T-patterns of attack actions between the first half and the second half.

Method

The present study exploits the observational methodology (Anguera, Blanco-Villaseñor, Hernández-Mendo, & Losada, 2011; Anguera, Blanco-Villaseñor, & Losada, 2001; Anguera y Hernández-Mendo, 2014), enhanced by the use of new technology. The observational methodology became more and more widespread within sport research (Torres, García-Mas, Palmer, y Cruz, 2008; Sautu Apellániz, Garay, y Hernández-Mendo, 2009; Serra Oliveses, González Villora y García López, 2011; Villarejo, Ortega Toro, y Palao Andrés, 2012; Martín, González, Cavalcanti, Chirosa, y Aguilar, 2013), because of its high flexibility and adaptability (Anguera, 2009).

The methods of analyzing performance in the game of soccer have evolved from the simple use of hand notation tracking of players’ movements on scale plans of pitches (Bloomfield, Jonsson, Polman, Houlahan, & O’Donoghue, 2005) to the current utilization of digital video recordings and computerized analyses (Borrie, Jonsson, & Magnusson, 2002). Digital recordings and computer analysis (e.g., Borrie, Jonsson, & Magnusson, 2002) have been largely used in sports research (e.g., Luo, Wu, & Hwang, 2003) and specifically in soccer (e.g., Camerino et al., 2012), because of the great advantages brought in terms of recording quality, measurement of time and capture of co-occurrences or diachrony.

Design

As Anguera and colleagues (2001; 2011) proposed, we consider three main criteria to give a taxonomic definition of the observational design, as applied to our study. The observational design was nomothetic (as opposed to idiographic, which refers to the number of subjects observed), since we observed different matches; punctual (as opposed to continuous, referring to the number of observations conducted on the same subject); multidimensional (as opposed unidimensional, referring to the number of levels of response observed).

The adoption of this N/P/M (nomothetic, punctual, multidimensional) design led to a series of decisions being made regarding the structure of the observation instrument, the type of data, data quality control and data analysis.

Participants

This study is part of a broader research project involving the analysis of all games played by a top club during the first leg (19 matches) of the Italian National League Championship (Serie A) over the 2012-2013 season (Anguera, Zurloni, Jonsson, Diana, & Elia, 2013). For this study, we chose to take in consideration three matches, randomly selected among those matches in which there was a positive change (from lose/tie in the first half to tie/win in the second half) in the result between first and second half.

Procedure

Coding instrument. We used Lince behavior coder (Gabín, Camerino, Anguera, & Castañer, 2012; Hernández-Mendo et al., 2014) to analyze and code the events of each match taken in consideration.

Lince software is consistent with the proposed observational design, since it is multidimensional in nature and structured around fixed, mixed and changing criteria (Castellano, Perea, Alday & Hernández-Mendo, 2008; Hernández-Mendo, López-López, Castellano, Morales-Sánchez, y Pastrana, 2012). Lince can also calculate Cohen’s kappa coefficient (Cohen, 1960, 1968) for all or some of the criteria by comparing two registered data files.

Observation instrument. The observation instrument combines field format and category systems (Anguera, Magnusson, & Jonsson, 2007). The fixed criteria are entered at the
beginning of the match, while the mixed criteria apply every time there is a change in the score, number of players and between the first and the second half of the match. The changing criteria are coded throughout the whole match. Each of these criteria gives rise to respective category systems that fulfill the conditions of exhaustiveness and mutual exclusivity (E/ME).

In this study, we decided to focus on attack actions mainly because of shooting restrictions. Since our video data were obtained from the TV recordings of the matches, the focus of the video was always the active play, in particular the player in ball possession. Since defense tactics comprehend a certain amount of moving and changing positions without ball possession, they were only partially observable using a public TV shooting.

The dimensions considered in the present observation instrument correspond to the following criteria: lateral position, zone (see Figure 1), lateral passing, zone passing, recovery and loss, ball out of play (see Table 1).

We defined an attack action as an action that brings the ball in the ultra-offensive zone and can end with: Goal (G), Non-Goal (NG), Permanent Loss (PL); this includes penalties, corner kicks, free kicks in ultra-offensive zone and throw-ins in the ultra-offensive zone.

Each attack action was coded starting from the first pass that crosses the offensive line. The same happens in case the ball gets to the ultra-offensive zone thanks to a free kick or a throw-in conceded to the observed team.

Analysis instrument. Theme 6 beta was used to analyze datasets. This software detects the temporal and sequential structure of datasets, revealing repeated patterns that may regularly or irregularly occur within a period of observation (T-patterns). A T-pattern is essentially a combination of events where the events occur in the same order with the consecutive time distances between consecutive pattern components remaining relatively invariant (Magnusson, 2005).

Results

Data quality

Two observers used Lince software to code the three selected matches. The same software calculated Cohen’s kappa coefficient for all the criteria by comparing the two registered data files. The values ranged between 0.82 and 0.88, which provides a satisfactory guarantee of data quality. However, when particular disagreements were identified, the specific cases were discussed and agreed on by the two coders.

T-patterns in each match

The individual datasets were analyzed ($p<0.005$; minimum occurrence = 3) with Theme in order to search for the most relevant T-patterns in each match and condition (first/second half). T-patterns were detected applying the default search parameters in Theme.

The number of patterns is increased between the first and the second half, in all the matches selected (Table 1). No differences were revealed considering length mean and level mean.

| Table 1. Total number of T-patterns (PatOccs) and number of different T-patterns (PatDiff) in first and second half. |
|---------------------------------|---------|---------|---------|---------|
|                                 | First half | Second half |
|                                 | PatOccs | PatDiff | PatOccs | PatDiff |
| First match                     | 9       | 3       | 30      | 9       |
| Second match                    | 51      | 15      | 85      | 23      |
| Third match                     | 21      | 6       | 52      | 14      |

Figure 1 and 2 shows two of the more complex T-patterns selected, respectively, from the first half and the second half of the same match. In particular, the T-pattern in the first half was detected for three times near the end of the half. It shows an attack action that starts with a pass from the offensive central zone to the offensive right zone, followed by a player that keeps ball in the ultra-offensive right zone, and ends with a pass from the ultra-offensive right zone to the ultra-offensive central zone (Figure 1).
Figure 1. A T-pattern detected for three times in the first half. It shows an attack action that starts with a pass from the offensive central zone to the offensive right zone (o, ce, op, tp), followed by a player that keeps ball in the ultra-offensive right zone (uo, ri, kb), and ends with a pass from the ultra-offensive right zone to the ultra-offensive central zone (uo, ri, uop, cep).

The T-pattern in the second half was detected as occurring three times. It describes a corner kick from the ultra-offensive right zone, followed by a pass from ultra-offensive right zone to the ultra-offensive central zone, and ends with a shot from the ultra-offensive zone (Figure 2).

Figure 2. A T-pattern detected for three times in the second half. It shows a corner kick from the ultra-offensive right zone (uo, ri, ck), followed by a pass from ultra-offensive right zone to the ultra-offensive central zone (uo, ri, uop, cep), and ends with a shot from the ultra-offensive zone (uo, ce, ng).
Comparison between first and second half

A second analysis was performed ($p<0.001$; minimum occurrence = 3) on the 100% of the datasets, in order to identify the patterns that allow us to distinguish the organization of play between first and second half.

The number of patterns (101/159), their length mean (3.17/4.24) and level mean (1.95/2.49) is increased between the first and the second half.

When we selected only those T-patterns where a shot (ng) or a goal (g) were present, similar distribution of the number of patterns (5/9), their length mean (2.4/3.78) and level mean (1.4/2.67) was detected. They all increased between the first and the second half.

Only 5 common T-patterns were detected in the first half. The most complex T-pattern (see Figure 3) was detected once in each match. It describes an attack action that starts with a player who keeps ball in the offensive-right zone, followed by a momentary loss in the ultra-offensive zone, and ends with a shot in the ultra-offensive central zone.

Figure 3. The most complex T-pattern from first half-related datasets. It shows an attack action that starts with a player who keeps ball in the offensive-right zone (o, ri, kb), followed by a momentary loss in the ultra-offensive zone (uo, ce, ml), and ends with a shot in the ultra-offensive central zone (uo, ce, ng).

9 common T-patterns were detected in the second half. Figure 4 and 5 show two of the most complex T-pattern. They were both detected one time in each match, and they both ended with a goal. Figure 4 describes an attack action that starts with a player who keeps ball in the offensive-central zone, followed by a pass from offensive central to ultra-offensive central zone, again a player keeps ball in offensive central zone, a pass from offensive central to offensive left zone, a permanent loss in ultra-offensive central, and ends with a goal from ultra-offensive central zone.
Figure 4. The most complex T-pattern from second half. It describes an attack action that starts with a player who keeps ball in the offensive central zone (o, ce, kb), followed by a pass from offensive central to ultra-offensive central zone (o, ce, uop, cep), again a player keeps ball in offensive central zone (o, ce, kb), a pass from offensive central to offensive left zone (o, ce, op, lp), a permanent loss in ultra-offensive central (uo, ce, pl), and ends with a goal from ultra-offensive central zone (uo, ce, g).

T-pattern in figure 5 describes an attack action that starts with a recovery in offensive central zone, followed by a pass from ultra-offensive left to ultra-offensive central zone, a pass from offensive right to ultra-offensive central zone, and ends with a goal from ultra-offensive central zone.

Figure 5. Another complex T-pattern from second half. It describes an attack action that starts with a recovery in offensive central zone (o, ce, t), followed by a pass from ultra-offensive left to ultra-offensive central zone (uo, le, uop, cep), a pass from offensive right to ultra-offensive central zone (o, ri, uop, cep), and ends with a goal from ultra-offensive central zone (uo, ce, g).
Discussion and conclusions

Data from the first analysis have shown a remarkable difference between first and second half in terms of game strategies. Comparing T-patterns’ structure and distribution between won and lost matches, striking differences were found. The total number of pattern occurrences and the number of different T-patterns detected was greater in the second half than in the first one, whereas the number of events coded per game was similar. This suggests that, in the second half, the team is more likely to continue using the same strategy.

Moreover, T-patterns of attack actions in the second half frequently ended with a shot, whereas there were no shots included in T-patterns of the first half. This result suggests that, as the result changes between first and second half, the game strategies that are repeated in all matches lead to effectively realize the amount of effort created. Such difference was exacerbated comparing common patterns of first halves and second halves: while in the formers no goals were detected within the T-patterns, the latters showed the presence of goals very frequently.

The results show that T-pattern analysis is an effective tool that supports research in sport performance analysis. Our preliminary data need to be deepened and confirmed by other analyses: a) through observation and comparison between all the matches of the first and second leg (the next 19 games of the championship); b) considering factors other than the result of the matches, such as the field factor (comparison between home and away matches); c) extending the observation not only to attack actions, but also to teams’ defensive behavior. However, the third point requires having ad hoc shootings of the games more than the TV ones, which focus on the ball possession and not on the whole team’s movements.

Nevertheless, our results highlight the potential for T-pattern analysis to make a significant contribution to the study of soccer dynamics. This type of analysis could provide the coach and the soccer teams’ staff with useful information that current analytical methods may not detect or overlook. This could help increasing the team’s performance at different levels.

This results also point towards the need to investigate the potential link between temporal structure detection and coach observations in soccer performance. This approach could assist in obtaining a greater understanding of coach knowledge construction.

Practical applications

The T-pattern methodology implemented in Theme software can help improving a team’s performance on different levels. Spontaneous and/or unaware tactics exhibited by the team can be detected through t-patterns, for one; on the other hand, it can be used to verify the application of tactics that were previously prepared and expected by the coaching team during training.

It also allows to identify recurrent behaviors related to effective or ineffective action plays, as well as analyzing the opposing team’s play and prepare countermeasures using what emerged through patterns, allowing to arrange (and rearrange) training programs according to what is detected by the analyses.

Theme and t-patterns mean the opportunity to use a tool that gives an alternative or additional measure of team performance, with respect to the commonly used statistics.

Acknowledgments: This research was supported by Spanish government project “Observación de la interacción en deporte y actividad física: Avances técnicos y metodológicos en registros automatizados cualitativos-quantitativos” (Secretaría de Estado de Investigación, Desarrollo e Innovación del Ministerio de Educación y Ciencia, Grant DIP2012-32124). We gratefully acknowledge the support of the Catalan government project Grup de recerca i innovació en dissenys (GRID). Tecnología i aplicació multimedia i digital als dissenys observacionals (Grant number 2014 SGR 971).

References

tamiento observado del árbitro de voleibol y su percepción por parte de las jugadoras: Una adaptación preliminar del CBAS. *Cuadernos de Psicología del Deporte, 8*(2), 5-18.


