Match & Player Analysis in Soccer: Computer Coding and Analytic Possibilities

Article *in* International Journal of Computer Science in Sport · January 2003

CITATIONS
3

READS
1,887

8 authors, including:

Angel Blanco Villaseñor
University of Barcelona
59 PUBLICATIONS **2,190** CITATIONS
SEE PROFILE

José Luis Losada
University of Barcelona
73 PUBLICATIONS **1,258** CITATIONS
SEE PROFILE

Antonio Hernández Mendo
University of Malaga
260 PUBLICATIONS **3,545** CITATIONS
SEE PROFILE

Gudberg Konrad Jonsson
University of Iceland
95 PUBLICATIONS **1,598** CITATIONS
SEE PROFILE

Match & Player Analysis in Soccer: Computer Coding and Analytic Possibilities

M. Teresa Anguera-Argilaga[1], Ángel Blanco-Villaseñor[1], José Luis Losada-López[1], Toni Ardá-Suárez[2], Oleguer Camerino-Foguet[3], Julen Castellano-Paulis[4], Antonio Hernández-Mendo[5] & Gudberg Jonsson[6]

[1]University of Barcelona, Barcelona, Spain, [2]INEF-Galicia, A Coruña, Spain,[3]INEF-Centre de Lleida, Lleida, Spain,[4]IVEF, Vitoria, Spain,[5]University of Málaga, Málaga, Spain,[6]Iceland University, Reykjavík, Iceland

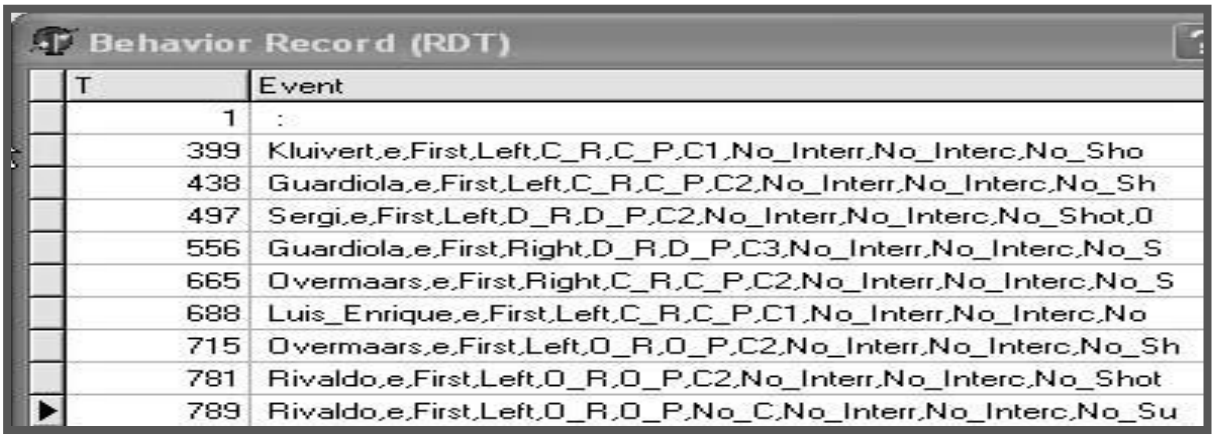
In team sports, especially in soccer, there is a certain complexity in the interaction between variables (players), variables of execution (displacements, lacks that will give place to regulation interruptions, throw, ball losses, recoveries, passes, etc.), and contextual variables (area of reception of the ball, area of pass of the ball, duration of the actions, etc.).

In order to achieve a systematized game analysis, we propose an instrument (SOF-3) that combines a structure of field formats with systems of categories, and it's substantially different that SOF-1 (Anguera, Blanco, Losada, Ardá, Camerino, Castellano y Hernández Mendo, in press; Anguera & Jonsson, 2002). The approaches of the instrument are: Times of the party, player, lateral spaces, area of reception of the ball, pass area, displacement, interruption, interception, shot, duration of the action, and marker. Each one of them, to the player's exception, gives place to a system of categories that, logically, will be exhaustive and mutually excluding. Also, it will record type of competition, pitch position, and the number of order of the plays of each one of the two teams, with the purpose of having a great database coming from different seasons (complete seasons) and from different teams. The code will be binary (0/1), and it will be carried out by computer, through software Excel (Figure 1), and, also, through software ThemeCoder (Figure 2).

Spell 2000-2001
Game Real Madrid - Barcelona

COMP		PER		TEAM				PLAY	DORS.	LATERAL		AREA RECEP.						AREA PASS						CONTACT					INTERRU			INTERCEP			SHOT			TIME	DUR.	MARK
L	H	1	2	O	C	TM	Inob			D	I	UD	D	C	O	UO		UD	D	C	O	UO		1	2	3	4	ØC	D	F	ØIR	P	R	ØINR	SE	CE	ØT			
L	H	1	0	1	0	0	0	1	Kluivert	0	1	0	0	1	0	0		0	0	1	0	0		1	0	0	0	0	0	0	1	0	0	1	0	0	1	0:15		0-0
L	H	1	0	1	0	0	0	1	Guardiola	0	1	0	0	1	0	0		0	0	1	0	0		0	1	0	0	0	0	0	1	0	0	1	0	0	1	0:17	0:02	0-0
L	H	1	0	1	0	0	0	1	Sergi	0	1	0	1	0	0	0		0	1	0	0	0		0	1	0	0	0	0	0	1	0	0	1	0	0	1	0:22	0:05	0-0
L	H	1	0	1	0	0	0	1	Guardiola	1	0	0	1	0	0	0		0	1	0	0	0		0	0	1	0	0	0	0	1	0	0	1	0	0	1	0:26	0:04	0-0
L	H	1	0	1	0	0	0	1	Overmaars	1	0	0	0	1	0	0		0	0	1	0	0		0	1	0	0	0	0	0	1	0	0	1	0	0	1	0:27	0:01	0-0
L	H	1	0	1	0	0	0	1	Luis_Enrique	0	1	0	0	1	0	0		0	0	1	0	0		1	0	0	0	0	0	0	1	0	0	1	0	0	1	0:28	0:01	0-0
L	H	1	0	1	0	0	0	1	Overmaars	0	1	0	0	0	1	0		0	0	0	1	0		0	1	0	0	0	0	0	1	0	0	1	0	0	1	0:30	0:02	0-0
L	H	1	0	1	0	0	0	1	Rivaldo	0	1	0	0	0	1	0		0	0	0	1	0		0	1	0	0	0	0	0	1	0	0	1	0	0	1	0:31	0:01	0-0
L	H	1	0	1	0	0	0	1	Rivaldo	0	1	0	0	0	1	0		0	0	0	1	0		0	0	0	0	1	0	0	1	0	0	1	1	0	0	0:32	0:01	0-0

Figure 1. Excel record with SOF-3 instrument.



Behavior Record (RDT)

T	Event
1	:
399	Kluivert,e,First,Left,C_R,C_P,C1,No_Interr,No_Interc,No_Sho
438	Guardiola,e,First,Left,C_R,C_P,C2,No_Interr,No_Interc,No_Sh
497	Sergi,e,First,Left,D_R,D_P,C2,No_Interr,No_Interc,No_Shot,0
556	Guardiola,e,First,Right,D_R,D_P,C3,No_Interr,No_Interc,No_S
665	Overmaars,e,First,Right,C_R,C_P,C2,No_Interr,No_Interc,No_S
688	Luis_Enrique,e,First,Left,C_R,C_P,C1,No_Interr,No_Interc,No
715	Overmaars,e,First,Left,O_R,O_P,C2,No_Interr,No_Interc,No_Sh
781	Rivaldo,e,First,Left,O_R,O_P,C2,No_Interr,No_Interc,No_Shot
789	Rivaldo,e,First,Left,O_R,O_P,No_C,No_Interr,No_Interc,No_Su

Figure 2. ThemeCoder record with SOF-3 instrument.

About this instrument (SOF-3), we have built a codifying manual, with some sintactic rules.

The wealth of obtained information resides in the nature of the data, and in that facilitates, by means of univariant and multivariant analysis, the search of diverse relationships among the suitable approaches of the instrument. The main techniques of analysis of data that will be used, will depend on the function of the characteristics of each one of the variable (approaches of the instrument), and they will facilitate the decisions taking, keeping in mind all the available information in the analysis situation.

In this study we apply the loglineal modelization that describes symmetrical relations between the variables. It has been used to develop to this technique procedure CATMOD (SAS Institute, 1989). The table of the analysis of variance by maximum likelihood contributes information for each one of the effects including of the analyzed model. The probability reason is an indicator of the goodfitnessness of adjustment, and the statistical Wald indicates the significant and nonsignificant effects that they contribute to this adjustment. Once determined the adjustment of the raised model (in our case triple association), they have estimated their λ parameters, or, alternatively, they calculate odds ratio of the different effects that participate in the model to know in each one them the degree intensity that is pronounced in the relations between the diverse variables.

The graphical representation of the loglineal models uses several methods to study the pattern of the association between the rows and columns in the contingency tables 2x2 (Cohen, 1980, Friendly, 1991, Snee, 1974). The technique of the *mosaic* shown by Hartigan and Kleiner (1981), can be used to represent the residual values of any loglineal model. In the mosaic each cell of the table is represented by a rectangle or area, but each one of them is proportional to the frequency of the cell. The mosaic is built dividing a square in a vertical sense, representing a variable (*Balloon Contact* category), and it will be divided in an horizontal sense representing the other variable (*Interruption* category).

Another alternative of graphical representation consists of circumferences that represent the residual values fit based on the intensity (Leverage) of the effects of model (*D value* of Cooks).

References

Anguera, M.T., Blanco, A., Losada, J.L., Ardá, T., Camerino, O., Castellano, J. y Hernández Mendo, A. (in press). Instrumento de codificación y registro de la acción de juego en fútbol (SOF-1). *Revista Digital de Alto Rendimiento en Fútbol*.

Anguera, M.T. & Jonsson, G.K. (2002, June). *Detection of real-time patterns in sports: Interactions in football*. Third Meeting of the European Research Group on *"Methodology for the analysis of social interaction"*. Milan: Catholic University of Milan.

Cohen, A. (1980). On the graphical display of the significant components in a two-way contingency table. *Commun. Statist.-Theor. Metho. A9*, 1025-1041.

Friendly, M. (1991). *Mosaic displays for multi-way contingency tables*. York Univ.: Dept. of Psychology Reports, No.195.

Hartigan, J,A., Kleiner, B. (1981). Mosaics for contingency tables. En W.F. Eddy (Ed.), *Computer Sicience and Statistics: Proceedings of the 13 th Symposium on the Interface*. New York: Springer-Verlag

SAS Institute, Inc. (1989). *SAS/STAT User Guide Versión 8.2* (Vol. 2). Cary, North Carolina: SAS Institute Inc.

Snee, R.D. (1974). Graphical display of two.way contingency tables. *The American Statisticican*, *28*, 9-12.