

Football and data mining: a statistical system for a “real-time” monitoring of players’ performance in the European Championships

Furio Camillo (Dipartimento di Scienze Statistiche Università di Bologna)

Stefano Sforza (IDS Information Delivey Systems Srl)

Abstract

The paper aims to present various statistical models for the monitoring in real time of what happens in a football match. In particular, we discuss an evolution in terms of data-mining of the Digital Soccer System which has been used from some RAI journalists to discuss and to give a commentary on the games of the last European Football Championships.

The research, and the experimentation of the used statistical models, have been entirely carried out in SAS environment, beginning from Software specific of “real-time survey” of what happens during the games, and of geo-reference, on the field, of all the noticeable athletic gestures according to previously defined technical codes. We try to show how it has been possible, working within the SAS environment and moving with a “Data-Mining approach” to implement models of synthetic appraisal and monitoring of the performance of a single soccer player during one sport competition like Euro2000.

Some considerations about the evaluation of the performance of a football player

One of the most read column on the press, is without doubt, that one of the reports. The synthesis, expressed with a simple number, of a complex performance as that one of a soccer player

All the newspapers dedicate, the day after the contests, wide space to the free interpretations of the own journalists which became classification of merit written up simply calculating the medium-mark of the football player. But, as all the subjective appraisals also those reports suffer of the problems concerning measurements in which the human contribution is preponderant. In fact, the appraisal of the rendering of a soccer team is a very complex task. Complexity derives substantially from two factors:

- 1- Football is a game in which many random elements are present.
- 2- Differently from many other sports, Football is based on the rarity of the determining events, first of all the goal: rarely we can see more than 4 or 5 goal scored in one match.

These two considerations imply some successive reflections. The trainer during the game succeeds in monitoring the team seen as a collective of athletes that must carry out some operations, with sure rhythms and one sure continuity. Not always it is possible to concentrate to the objective rendering of a single player: the exact amount of actions or technical gestures that the player, during the contest (or the

various fractions of the contest), carries out in field. Moreover the variety of the contexts in which the player can be could be too large for a correct analysis of the informations available. The conditionings that the module, the environmental factors, the adversaries or other factors can determine are remarkable and the use of statistical models can concur the decomposition of the total performance of one team in the various members: performance of the single one, performance of the adversary, outlines of game, environmental conditions or contingent (insufficient physical shape, fatigue in the physical recoveries, etc).

The DigitalSoccer Project

A precise and shared standardization of the elements of the game and a wide collection of data are presupposed for an “objective evaluation” of the game. The DigitalSoccer System, implemented by the Society DigitalSoccer Project, was born with both requirement. Thanks to the collaboration with a pool of trainers of first level (Sacchi, Lippi, Zeman, Lucescu, Ancelotti, etc) it has been defined a common language, used from 1994 also from the Technical Field of the Federation. At the same time it has been created a system of survey which is able, in real time, to analyse and to process all the collected technical and tactical informations.

In particular, 4 situations of a game have been standardized:

- the action of the player who is in ball possession: it concerns the technical gesture (cross, long passage, shooting, etc) and its contextualization (according to the outcome of the action it can become lost ball, useful passage, assist, assist winning, network).
- the action of the player who " dictates " the passage: it is a movement without ball that is defined in relation to the position of the ball bearer (superimposition, inner cut, external cut, etc).
- the action of the player who wants to recover the ball: the recovery is distinguished during the transmission of the football (interception, advance payment, elevation), and the recovery with the adversary guiding the ball (contrast and doubling).
- the action of the whole team: abilities are the defensive movements of the collective (to keep the adversaries in offside or to force them to mistake), and all the endured actions (such as fouls, cross, shooting, etc).

Cataloguing all the actions of game, for the team as a whole and for all the single players, in the database it is possible to have complete technical-tactics information about the development of the entire contest. Through the DigitalSoccer Software it is possible to calculate the medium data of every player and, confronting them with those of the role of reference, it is possible to express in automatic way a judgement on every single ability. In this way you can find out that, as an example, that the player *X* has one detached attitude to the dribbling (he tries double number regarding the average of role), but he is hardly sufficient in the effectiveness of the same ones (he succeeds little more than one on four). Or it can be discovered that the player *Y* is optimal in the ability to gain the contrasts, but only sufficient in shooting.

The IVG System and the SAS System

In order to synthetize these single information in a total judgement on the performance of the football player the most pertinent indexes, for everyone of the 4 fundamental roles, are selected. The summary obtained by processing these weighted means determines the **IVG Index** (General Valuation Index) that, in the respect of the complexity of the analysis, is expressed in thirtieth. It is probable that trainers, for separated games and for every single player, have some expectations on the single's performance. Such expectations are, however, difficultly quantifiable in objective terms, that is making tightened reference to the number of actions or specific technical gestures attended in a precise zone of the field from the single player and in one specific phase of the match. In other words, it is difficult to estimate before the game how many, as an example, the expected passages or how many assist or how many interceptions in the midfield or in the defensive zone from the single player in that game, should be. The IVG of Digital Soccer tries to resolve in automatic way such problems defining like "expected" that performance that is in tune with the historical average of the role of the single player. The IVG is therefore a statistical index of probability for a specific technical event that can happen in the field. The index is based on the information available about the past and therefore on the historical memory contained in the database. One of the pluses of the IVG System is that it synthetizes the performance of a single player beginning from one great amount of determining facts. In such sense system IVG could, in a not too much far away future, represent a tool of analysis with which the objective rendering of the single ones, in real time and with a terminal on the bench, is monitored. As we said, for its statistical characteristic the IVG System can be meant like a statistical pointer for the estimate of the probability for a specific technical event to happen, but also it can be used to estimate the total probability of victory for one team. The calculation in real time of the IVG gives the possibility to dynamically

monitor the synthetic performance of the single player. That is rendered from a statistical model that, minute by minute, estimates, on the base of 19 found technical actions/gesture in field, which it will be the final performance of the player and therefore its specific possible IVG at the end of the game. It is a forecast model that, in a precise moment of the game, answers to the question: "*which would his final IVG be, if the player continued to carry out the same size of up to now carried out job?*". Obviously the model holds account of some important variables such as passing of the time, that is of the fact that the physical rendering of a player diminishes with passing of the minutes and that, therefore, as an example, a player entered in field to the 75° tiny will have greater probabilities to realise useful technical actions to the measure of his IVG. On the other hand, the fact that it is an objective system, must, above all, be meant like a limit of the system itself. In fact it will not be never possible to estimate some not tangible aspects like competitive commitments of the single player or the his ability to direct the companions in some crucial situations. These last ones are appraisals than only a trainer can succeed to carry out by means of his own experience and his own technical and human sensibility.



Filiale di Padova
Via della Croce Rossa, 112
35129 Padova
049 8994511



UniBologna

INFORMATION
DELIVERY
SYSTEMS
LabUnit



Furio Camillo

Stefano Sforza

Football and data mining: a statistical system for a "real time" monitoring players' performance in European Championships



SeUGI 19

www.sas.com



SeUGI 19

www.sas.com

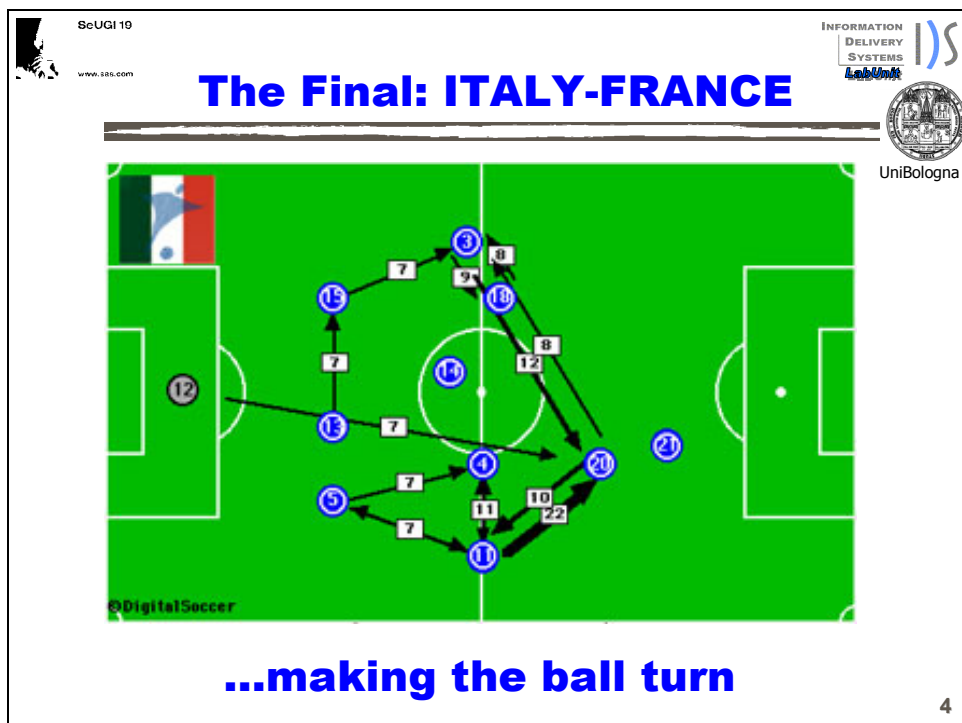
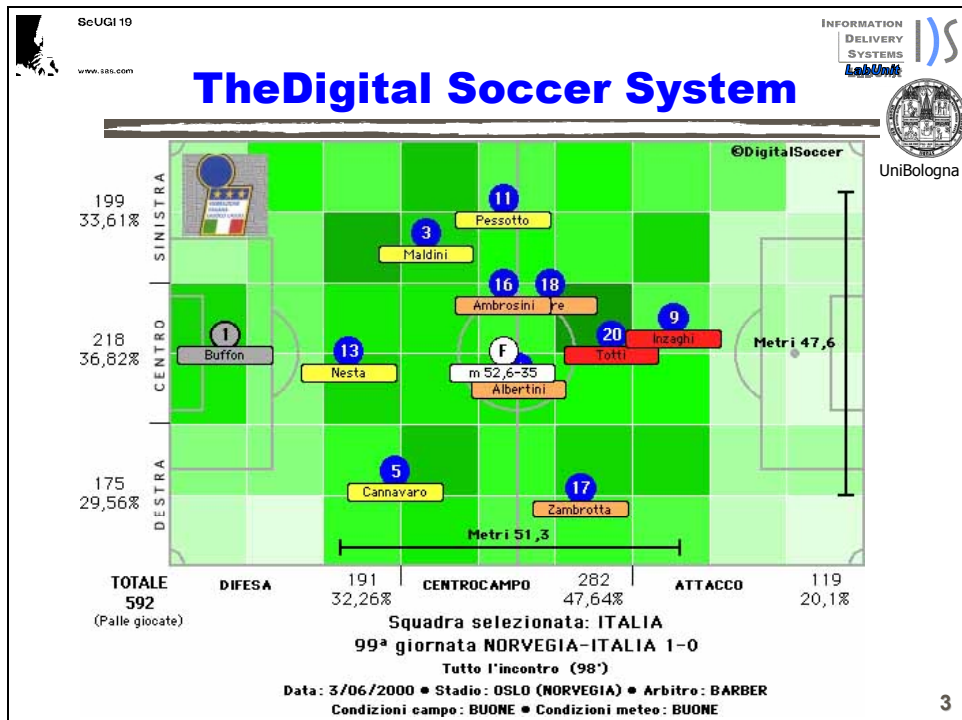
Introduction

INFORMATION
DELIVERY
SYSTEMS
LabUnit



UniBologna

- **Statistical models for the monitoring in real time of a football match**
- **Problems concerning the reliability of the evaluation of the performance of a single player by means of the human eye**
- **Interactions of gesture and events**
- **Rarity of the events**
- **Digital Soccer Project implemented a method to analyse these events**
- **In collaboration with "the world of football": Lucescu, Zeman, Ancelotti, Sacchi, Zaccheroni, Malesani and others**
- **At Euro2000 the IVG System has been implemented**



SeUGI 19

www.sas.com

INFORMATION
DELIVERY
SYSTEMS

LabBook

UniBologna

The Final: ITALY-FRANCE



...looking for the playmaker: Zidane

5

SeUGI 19

www.sas.com

INFORMATION
DELIVERY
SYSTEMS

LabBook

UniBologna

Four typical situations

- **The action of the player who is in ball possession:** it concerns the technical gesture (cross, long passage, shooting, etc) and its contextualization (according to the outcome of the action it can become lost ball, useful passage, assist, winning assist, goal).
- **The action of the player who " dictates " the passage:** it is a movement without ball that is defined in relation to the position of the ball bearer (superimposition, inner cut, external cut, etc).
- **The action of the player who wants to recover the ball:** the recovery is distinguished during the transmission of the football (interception, elevation), and the recovery with the adversary guiding the ball (contrast and doubling).
- **The action of the whole team:** abilities are the defensive movements of the collective (to keep the adversaries in offside or to force them to mistake), and all the endured actions (such as fouls, cross, shooting, etc).

6



DSP, the IVG and the research

- Some recent development for statistical methodology have concurred to implement a method to estimate a simple synthetic index of performance
- The key concept is "The EXPECTED VALUE"
- By means of system DSP we can estimate the expected value for every player basing on the historical average of its role
- In this way the observer (trainer or media operator) can obtain a probabilistically correct estimate of the performance of every player, supported by objective elements of evaluation