A Novel Decision Support Model to Discover the Interesting Pattern in Football Match

Bencheng Chai[1,2], Xu Xu[3,4]
[1] School of Computer Science and Information Technology, Zhejiang Wanli University, Ningbo, Zhejiang 315100, P.R.China
[2] School of Electronic and Information, Tongji University, Shanghai, 201804, P.R.China
[3] School of Economics and Management, Tongji University, Shanghai, 200092, P.R.China
[4] Laboratory of Intelligence Decision and Decision Support, Tongji University, Shanghai, 201804, P.R.China
chaicbc@163.com, xuxusuccess5616@yahoo.com

Abstract

This paper utilizes a novel decision support model to study the pattern in the football match, and to find the interesting association rule and cooperation custom. Data mining technique, serial data, pass time and ball possession are employed to build the decision model. The proposed model is not only expatiated in the article, but also is proved by the experiments. In short, the model provides a feasible route to guide the football coach and athletes to establish effective system against the opponent tactics.

1. Introduction

Much of opponent information (game video, training video, etc) in football sport has often been considered in the traditional training by coach and athlete. However, the training video is difficult to get, and the achieved training videos are only some of those training videos. So the reference value is not feasible. The study of match video often locates the concrete match video or some of them. The pattern or association rules in the match videos are not often considered in the scientific research. Recently, with the development of technology, more and more researchers utilized computer science methods, management approach, mathematics model to predict or discover some useful pattern and rules in sports fields [1-2].

Football is a competitive sport with two teams of 11 players. The winning team must score more goals than the other team during the game. Most research [3-4] of football sports focused on statics and the effective of attack. A Markov process model is used to estimate transition probabilities by means of the maximum likelihood method [5]. The basic characteristics of football mean that it can be analyzed just like any productive activity [1].The concept is employed in our research. The winning team is good cooperation team. The cooperation custom and pass pattern will be very important to the winner part. But the other part can prevent the cooperation and effective pass ball by the discovery of cooperation custom and pass pattern.

The main aims of this article are as follows:

- build the decision support model
- Discover the cooperation custom and pass pattern
- Establish effective system against the opponent tactics

The novel decision support model is consisted of data mining technique, serial data, pass time and ball possession. These contents will be discussed in later section.

2. Decision support model

The model includes the following data resources:

- Serial data, pass time
- Serial data, ball possession
- serial data, pass time, and ball possession

This paper combines serial data, pass time, and ball possession to establish the transaction database.Then data mining is used to find the intertesting pattern.

2.1. Serial Data

Football is a competitive sport with two teams of 11 players. So there are 22 players on the playground. In the proposed novel model, A_i denotes the number of

978-0-7695-3735-1/09 $25.00 © 2009 IEEE
DOI 10.1109/FSKD.2009.676

one part players , where $i \in [1\ldots,11]$, B_j denotes the number of the other part players, where $j \in [2\ldots,11]$. The traditional approach only considers the movement of players. The proposed model considers not only the movement of players, but also the movement of the football. The latter part is the key point in the model.

Definition 1. Let Y_k be the series, i.e., Y_k is the combination of A_i and B_j . Where k satisfies $k \in N$. The last element of Y_k is different to the former element of Y_k (It includes that the football is kicked out).

Suppose Y_1, Y_2, where $Y_1=A_1A_2A_4A_5$, $Y_2=A_1A_2A_4A_5B_1$. According to Definition 1, Obviously, Y_1 does not belong to Y_k, while Y_2 belongs to Y_k. Y_k can express possession and dispossession directly.

2.2. Pass time and ball possession

In order to discover the cooperation custom and pass pattern effectively, pass time and ball possession are employed in the novel model.

Definition 2. Let P_u denotes pass time, P denotes the average value of P_u, where $u \in N$, and $u \in [1,v]$.

$$P = 1/v \sum_{u=1}^{v} P_u \quad (1)$$

Definition 3. Let B_m denotes ball possession, B denotes the average value of B_m, where $m \in [1,n]$, $m \in N$.

$$B= 1/n \sum_{m=1}^{n} B_m \quad (2)$$

2.3. Data mining

Data mining combines the statistic and artificial intelligence to find out the rules that are contained in the data, letters, and figures and soon by sorting and analyzing [6]. There are many methods of data mining including classification, estimation, predication, clustering and affinity grouping. Among these, affinity grouping can discover the high frequency pattern and find which things appear frequently and simultaneously. A novel study on the revising and launching automobile type is based on decision support and business intelligence system [7]. In order to protect the privacy of patents completely, Xu xu and ect. employ intuitionistic fuzzy set, α-cuts, and Apriori algorithm to discovery the relative pattern of a disease under an anonymous way [8].

The necessity of extraction of meaningful information from huge time series databases (TSDB), which can be useful for decision maker, caused the development of methods of time series data mining (TSDM). The goal of data mining is the analysis of (often large) observational data sets to find unsuspected relationships and to summarize the data in novel ways that are both understandable and useful to the decision-maker [9-10]. Data mining combines the statistic and artificial intelligence to find out the rules that are contained in the data, letters, and figures [11]. The following list contains traditional time series data mining tasks:

- Segmentation: Spilt a time series into a number of "meaningful" segments.
- Clustering: Find natural groups of time series or time series patterns.
- Classification: Assign given time series or time series pattern to one of several predefined classes.
- Indexing: Realize efficient execution of queries.
- Anomaly Detection: Find surprising, unexpected patterns.
- Motif Discovery: Find frequently occurring patterns.
- Forecasting: Forecast time series values based on time series history or human expertise.
- Discovery of association rules: Find rules relating patterns in time series.

The time series is refer according to the observation record ordered set which arranged in time successively. A large volume of research in temporal data mining is focusing on discovering temporal rules from time-stamped data. A model of temporal pattern discovery from multiple time series based on temporal logic is proposed [12]. Lucia Sacchi and etc. propose a new kind of temporal association rule and the related extraction algorithm; the learned rules involve complex temporal patterns in both their antecedent and consequent [13]. Based on data of the time series database, Chong Zhu and etc. present a new frequent time series patterns finding algorithm [14]. TSDM is also used in stock market analysis [15-16] .

TSDM is becoming increasingly important as a knowledge management tool where it is expected to reveal knowledge structures that can guide decision making in conditions of limited certainty. Human decision making in problems related with analysis of time series databases is usually based on perceptions like ''end of the day'', ''high temperature'', ''quickly increasing'', ''possible'', etc.Though many effective algorithms of TSDM have been developed, the integration of TSDM algorithms with human decision making procedures is still an open problem. I.Z. Batyrshin and L.B. Sheremetov defined the main characteristics of perception-based TSDM approach and they described different methods used to precisiate

a meaning of perceptions in TSDB domains and to compose them in complex patterns, rules or relations [17].

In this novel model, serial data are organized to form transaction database, and serial attached with pass time, ball possession to form time transaction database. The concept of TSDM is introduced in this model. Decision making procedure in problems related with analysis of time series databases often use ideas like several days, near future, high price, very quickly increasing, slowly oscillating, high associated behavior, etc. defined on a time domain, on a range of time series values, on the set of possibility or probability values, etc.

3. Model examination

In order to explicate the novel model, the data comes from the final match between Spain and Germany in European Cup 2008.The interesting pattern and rules of Spain team can be achieved from the novel architecture model. Adobe Premiere Pro.2.0 is employed to collect ball possession and pass time.

Table 1 shows the members of Spain team and Germany team.

Table 1 Player table (Spain and Germany)

Code	Player name	Team	Code	Player name	Team
A	Casillas	Spain	L	Lehmann	Germany
B	Marchena	Spain	M	Friedrich	Germany
C	*Puyol*	*Spain*	N	Schweinsteiger	Germany
D	Iniesta	Spain	O	Frings	Germany
E	Xavi Hernandez	Spain	P	Klose	Germany
F	Torres	Spain	Q	Ballack	Germany
G	Fabregas	Spain	R	Hitzlsperger	Germany
H	Capdevila	Spain	S	Lahm	Germany
I	Sergio Ramos	Spain	T	Mertesacker	Germany
J	Senna	Spain	U	Podolski	Germany
K	Silva	Spain			

Table 2 Serial data

Transaction	Transaction	Transaction
IEGL	IFO	EJEDJS
EDGL	IKFKS	EDEQ
ECHJKGKFO	ICJKEDT	KFT
GFT	EGEFDV	GKDT
EJKCBTGUHGO	CGFV	JDKS
AKFS	FEHEFEHL	
DFT	IGEGL	

This paper extracts 7200 serial data records to find the pattern, Table 2 shows some of them. As is mentioned above,IEGL means Ramos pass the ball to Xavi and Xavi to Fabregas and the pass is plucked by Lehmann of Germany.

Mark Way (Data mining tool) is employed to find the interesting pattern from the above data set.From the result of analysis, There are some intertesting and valuable pattern. For example, player J,C,E,G pass the football each other and the cooperation will be probably broken up by the opposite players T, O. Obviously, the pattern is useful to the coach and players.

Compared with the decision model of serial data, the yielded pattern and association rules are more concrete and diverse, so coach and players can choose the proper pattern and rules to adjust the tactics against the opponent.

4. Conclusion

The paper proposes a novel model to find the interesting pattern in football match. The model is examined by the experiments. TSDM and traditional data mining are all considered in this model. Some useful patterns are yielded from the experiments, and they can guide the football coach and players to establish effective system against the opponent tactics. The achieved pattern from the decision model can predict the cooperation custom and pass pattern effectively. So the decision model is scientific and feasible. Of course, there are some drawbacks in the decision model, for example, the relative error range of time of TSDM should be studied in the further research. In the future, more work and experiments will serve as motivation for efforts in the development of decision model, to make the use the model more effective, more useful.

5. Acknowledgements

Authors contribute equally in the article.

6. References

[1] Randall A. Gordon., “Attributional Style and Athletic Performance: Strategic Optimism and Defensive Pessimism”, *Psychology of Sports and Exercise*, 2008, pp. 336-350.

[2] S. Dobson and J. Goddard, "Persistence in sequences of football match results: A Monte Carlo analysis," European Journal of Operational Research, 2003, pp. 247-256

[3] Clemen, R. *Making Hard Decisions: An Introduction to Decision Analysis*, Wadsworth Publishing Company, 2006.

[4] Yong Zhou, Feng Liu. “Research of Robocup Pass Strategy Based on Improved Q-Learning”, *Computer Technology and Development*, 2008, pp.63-66.

[5] N. Hirotsu and M. Wright, "Using a Markov process model of an association football match to determine the optimal timing of substitution and tactical decisions," *Journal of the Operational Research Society,* vol. 53, pp. 88-96, 2002.

[6] Jiawei Han, Micheline Kamber. *Data Mining Concepts and Techniques*, Morgan Kaufmann Publishers, 2001.

[7] J. Lin and X. Xu, "A novel approach to find the revising and launching automobile type," *5th International Conference on Fuzzy Systems and Knowledge Discovery*, 2008, pp. 489-492.

[8] X. Xu, J. Lin, F. Wang, and Y. Jiao, "Utilize a novel approach to find the relative pattern of patients of a disease," *Proceedings of the World Congress on Intelligent Control and Automation*, 2008, pp. 4201-4205

[9] D, Hand, H. Manilla, P. Smyth. *Principles of Data Mining,* MIT Press, 2001.

[10] B.L. Bowerman, R.T. O’Connell. *Time series and Forecasting: An Applied Approach*, Duxbury Press, Mass, 2000.

[11] L. X. Du, X. Xu, Y. Cao, and J.Y. Li, “A novel approach to find the satisfaction pattern of customers in hotel management”, *Proceedings of the IEEE International Conference on Fuzzy Systems* ,2008, pp. 11-14

[12] Z. Chen, B.-R. Yang, F.-G. Zhou, L.-N. Li, and Y.-F. Zhao, "A new model for multiple time series based on data mining," *2008 International Symposium on Knowledge Acquisition and Modeling*, 2008, pp. 39-43

[13] L. Sacchi, C. Larizza, C. Combi, and R. Bellazzi, "Data mining with Temporal Abstractions: Learning rules from time series," *Data Mining and Knowledge Discovery,* 2007, pp. 217-247

[14] Z. Chong, Z. Xiangli, S. Jingguo, and H. Bin, "Algorithm for mining sequential pattern in time series data," *2009 International Conference on Communications and Mobile Computing*, 2009, pp. 258-262.

[15] C. Luo, Y. Zhao, L. Cao, Y. Ou, and L. Liu, "Outlier mining on multiple time series data in stock market," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2008, pp. 1010-1015.

[16] J. Yao and S. Kong, "The application of stream data time-series pattern reliance mining in stock market analysis," *2008 IEEE International Conference on Service Operations and Logistics, and Informatics*, 2008, pp. 159-163.

[17] I. Z. Batyrshin and L. B. Sheremetov, "Perception-based approach to time series data mining," *Applied Soft Computing Journal,* 2008, pp. 1211-1221.