

The analysis of official data for soccer World Cup 2002-2014

Choi, Hyongjun¹.

¹*Department of Physical Education, College of Physical Education, Dankook University, Yongin, South Korea.*

Introduction

The sport science has been investigated the enhancement of performance in different ways. Especially, the performance analysis of sport has been improved and applied its knowledge into the field of sports. The academic approaches to analyse the performances that the artificial intelligent(AI) techniques were newly concerned in past years. Among the AI techniques, the Self-Organised Map(SOM) is a way to visualise data including the maintaining of variables' features (Kohonen & Somervuo, 2002; Choi & Ko, 2010; Sohn & Choi, 2013). However, the visualisation of the official soccer World Cups' data has not been investigated yet, even though the official data has been developed and expended widely. This study, thus, keened to investigate methods using the SOM for the visualisation of data.

Methods

the samples were totally 512 cases such as home & away team's data and 20 variables from 2002 to 2014 soccer World Cups were selected and gathered into Microsoft Excel 2010. And a R programme with SOM package was used to design the Sammon's mapping and Self-organizing map projection. Sammon's mapping techniques were designed as 3-dimensional structure and a size of the Self-organizing map was [10 10]. Also, the Bland & Altman plot was used to determine the function of prediction in the SOM that 75% of data was used for the training and 25% of data was used for the prediction.

Results

After the training of Self-organising map, the quantisation error was 3.906 and the topographic error was 1.248. Consequently, the results of the study were as following; Firstly, the Sammon's mapping technique was unable to descript the details of data hardly when the size of data was huge. Secondly, 3-dimensional Sammon's mapping technique was able to visualize data easily with x-y, x-z and y-z coordination graphs. Thirdly, using Self-organizing map brought an advantage of visualisation for data process using comprehensive neurons which could apply variables' characteristics into the results.

Discussion & Conclusion

This study was to compare 2 methods of AI for the visualisation of the official soccer World Cups' data. Between the methods, the Sammon's mapping was usefully spotted data on the 3D graph, but it was hard to confirm a specific data on the chart if the amount of data was huge. On the other hand, the SOM method was efficiently used to visualise the data with the original features of variables, and it could also make clusters of the official data within the cluster analysis. Consequently, the official data is able to present the performances if the AI techniques were able to be used.

References

- Choi, H. J., & Ko, B. G. (2010). *New approaches of statistics in sports*. Paper presented at the Japanese Applied statistics conference 2010, National Institute of statistics and mathematics.
- Kohonen, T., & Somervuo, P. (2002). How to make large self-organizing maps for nonvectorial data. *Neural Networks*, 15(8-9), 945-952.
- Sohn, J., & Choi, H. (2013). Are golf-shots distinguished by power control? Or it is just individual differences? *International Journal of Performance Analysis of Sport (e)*, 13(2), 212-224.