Applied Soft Computing 14 (2014) 81–87

Contents lists available at ScienceDirect

Applied Soft Computing

journal homepage: www.elsevier.com/locate/asoc

Using meta-regression data mining to improve predictions of performance based on heart rate dynamics for Australian football

Herbert F. Jelinek[a,b], Andrei Kelarev[a,e], Dean J. Robinson[c], Andrew Stranieri[d], David J. Cornforth[e,*]

[a] Centre for Research in Complex Systems, School of Community Health, Charles Sturt University, P.O. Box 789, Albury, NSW 2640, Australia
[b] Biomedical Engineering, Khalifa University of Science, Technology and Research (KUSTAR), Abu Dhabi, United Arab Emirates
[c] Corio Bay Sports Medicine Centre, Geelong, VIC, Australia
[d] School of Science, Information Technology and Engineering (SITE), University of Ballarat, P.O. Box 663, Ballarat, VIC 3353, Australia
[e] Applied Informatics Research Group, School of Design, Communication and Information Technology, University of Newcastle, Callaghan, NSW 2308 Australia

ARTICLE INFO

Article history:
Received 14 March 2013
Received in revised form 5 August 2013
Accepted 5 August 2013
Available online 31 August 2013

Keywords:
Feature selection
Regression
Meta regression
Data mining
Heart rate dynamics
Australian football

ABSTRACT

This work investigates the effectiveness of using computer-based machine learning regression algorithms and meta-regression methods to predict performance data for Australian football players based on parameters collected during daily physiological tests. Three experiments are described. The first uses all available data with a variety of regression techniques. The second uses a subset of features selected from the available data using the Random Forest method. The third used meta-regression with the selected feature subset. Our experiments demonstrate that feature selection and meta-regression methods improve the accuracy of predictions for match performance of Australian football players based on daily data of medical tests, compared to regression methods alone. Meta-regression methods and feature selection were able to obtain performance prediction outcomes with significant correlation coefficients. The best results were obtained by the additive regression based on isotonic regression for a set of most influential features selected by Random Forest. This model was able to predict athlete performance data with a correlation coefficient of 0.86 ($p < 0.05$).

1. Introduction

This paper reports on the prediction of performance of elite athletes during games of Australian Rules Football (AFL). The prediction is based on a variety of measures derived from weekly ECG recordings, playing field dimension and temperature for the day of the game. The performance of players is expressed using variables derived from geographical positioning system (GPS data) of football players during the games. The success of sophisticated soft computing techniques to predict player performance during matches is demonstrated in this work. The outcomes are models that may provide the team coaches with a new tool for selection of players for any particular match based on physiological performance of players in preceding weeks. The coaches may also use such models to plan and implement training regimes that focus not only on the individual player but also on characteristics of the match that is being targeted. For example, coaches may be able optimise capacity for running at greater speed, which is associated with better and more ball control and therefore a better score.

The prediction of athlete performance follows a history of performance analysis of athletes, which has been the focus of research for decades with an emphasis on motor skills and other physical factors [1]. The application of the information sciences to sport started with statistical analysis and calculations based on biomechanical data [2]. Since then the concept of sports informatics has grown to include modelling and simulation, databases, soft computing, pattern recognition and visualisation. Recently the advent of light weight mobile technological devices such as heart rate monitors, global positioning system units and other sensor devices have led to increased opportunities for the partnering of Information Technology with Medical and Sports Science. A number of recent studies have explored the application of these technologies to a range of sports including rugby and soccer, in order to more accurately understand player performance, conditioning and recovery [3,4].

High-performance sport such as football requires athletes to be at their optimal performance for a substantial part of the year, with the rest of the year being taken up with player selection and preseason training. Major factors are mental exhaustion, physical injury and over-training. Therefore it is important to have information for

* Corresponding author. Tel.: +61 2 4985 4069.
E-mail address: David.Cornforth@Newcastle.edu.au (D.J. Cornforth).

1568-4946/$ – see front matter.
http://dx.doi.org/10.1016/j.asoc.2013.08.010

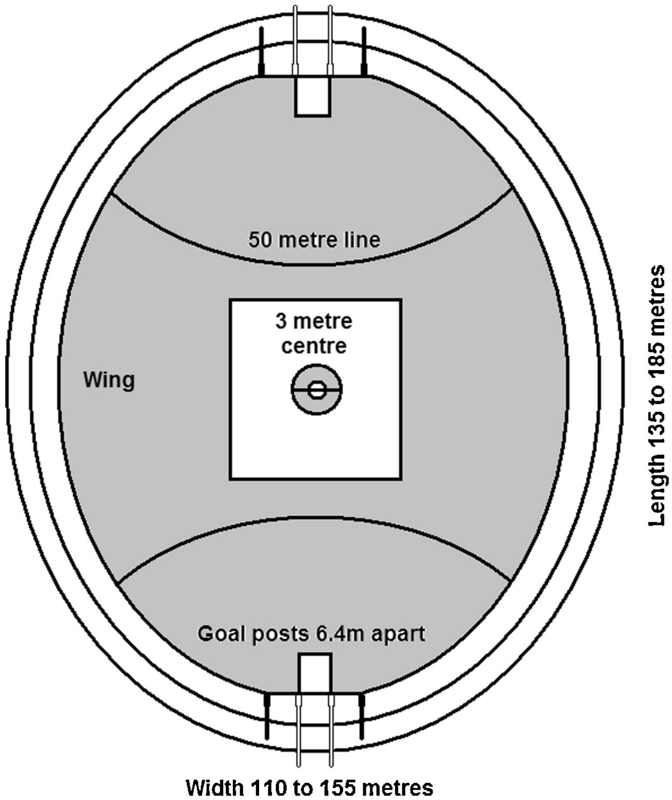

Fig. 1. Diagram of playing field for Australian football, showing oval shaped field and main dimensions.

the coach that reduces the risk of physical and mental exhaustion and injury, whilst maintaining a high performance level. Currently the football club at the centre of this study collects biometric data that is analysed either by sports exercise experts within the club or by third parties, neither of which have advanced knowledge of data processing or soft computing. There is a good opportunity for advanced regression techniques to contribute to novel advances in this area.

AFL is played with an oval ball on an oval field with a free flowing pace. Fig. 1 shows a diagram of the field along with the most important features, including the long and short dimension of the field. AFL stands in contrast to more structured styles of football, where a game is divided into well-defined phases, marked by a particular formation [5]. AFL is a fast moving game where speed and good ball handling and kicking skills determine the outcome. AFL adds extra variables to the analysis of fitness data because there is no standard size for the playing field, although the length is generally 135–185 m and the width 110–155 m. In this work the actual dimensions of the field for each game were recorded and used as one of the potential predictors of athlete performance. This is an important point because athletes have to make a different effort depending on the field that they find themselves playing on at any particular occasion. The prediction of athlete performance at any match is potentially confounded by this variable, so it should be included in regression models.

Player performance has also been reported to change throughout a playing season in the game of soccer [6]. This variability may be related to training status, perceived wellness and environmental conditions [7]. With respect to the latter, there have been reports of a greater incidence in injuries in football played in warmer and/or drier conditions [8]. Changes in core body temperature may also be a factor in performance but has only been studied under warmer

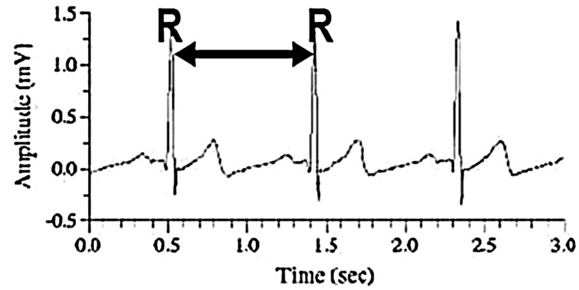

Fig. 2. Normal ECG recording showing RR interval.

conditions [9]. Our study therefore includes match temperatures as an additional variable.

There are numerous physiological variable measures in high performance sport. Heart rate is one known to be associated with fitness and performance. Heart rate variability (HRV) is based on data describing the time interval between successive heart beats. HRV is an indicator of the regulation of the heart [10]. A standard ECG signal is shown in Fig. 2. This type of signal has been extensively studied and the diagnostic value of the different features is well established. ECG features are referred to using letters. The large spike is referred to as the QRS complex, with R being the peak of the wave or fiducial point, while the smaller peak to the left is the P wave and to the right of the QRS is the T wave. The signal is degraded by the presence of noise, so that the most reliable feature that can be obtained from low quality recordings (and therefore the most easily obtained measurement) is the interval between successive R peaks, known as the RR interval or inverse of the heart rate. This can be expressed as millisecond intervals and varies considerably between individuals. The inverse being heart rate is typically 60–80 beats per minute.

The natural rhythm of the human heart is subject to variation that is believed to indicate the health of the cardiovascular system in that too much or too little variability between beats increases the risk of arrhythmia. RR intervals are obtained from the recorded ECG and subjected to further analysis through a variety of algorithms in order to yield variables with good discriminant power [10]. However, the RR interval can also be obtained from a simple and cheap chest strap device or incorporated into the playing jersey (garment), which makes it particularly suitable as a non-invasive test which allows data to be easily collected from players.

As the electrical system of the heart is a complex and adaptive system, variation between beats can also be used to assess the health of elite athletes. The response of the cardiac autonomic nervous system can be assessed non-invasively using a number of measures based on HRV [11], all of which can provide useful information regarding the functional adaptations to a given training stimulus. As heart rate increases due to exercise, HRV tend to decrease. This is known as the sympathetic effect and is manifest in lower frequency changes in heart rate. Conversely, during rest the heart rate slows and higher frequency variation predominates. This is known as the parasympathetic effect. High and low frequency domain parameters are obtained by fast Fourier transform from the RR intervals of the biosignal. The extent of variability in the frequency domain can therefore be used as an indicator of cardiac modulation by the autonomic nervous system and an indicator of general health of the athlete or level of stress.

HRV is rapidly becoming more accessible with the availability of wireless sensors at low cost. HRV has been conventionally analysed with time- and frequency-domain methods, which measure the overall magnitude of fluctuations around the mean or the magnitude of fluctuations in some predetermined frequencies. More recent nonlinear analysis has shown an increased sensitivity for