HIGH STAKES *horseracing*

soccer *financial* lottery
roulette *poker* blackjack
baccarat Backgammon
horse-racing *greyhound*
NFL *spread betting*

FORECASTING METHODS FOR HORSERACING

PETER MAY

HIGH STAKES

PETER MAY

FORECASTING METHODS FOR HORSERACING

HIGH STAKES

This edition published in 2002 by
High Stakes Publishing
21 Great Ormond St
London WC1N 3JB
T: 020 7430 1021
www.highstakes.co.uk

www.gamblingbooks.co.uk/publishing

ISBN 1-84344-002-4 Forecasting Methods for Horseracing

2 4 6 8 10 9 7 5 3 1

Printed and bound in Great Britain by Cox & Wyman

ACKNOWLEDGEMENTS

Firstly I would like to thank my friends and colleagues at Oxford Brookes University for their contributions, direct or indirect, to this book. I would especially like to thank Dr Nigel Crook, an expert in the field of artificial intelligence.

Secondly, I would like to express my gratitude to my family and racing friends for their help and contributions, as well as the proof readers John Taylor and Sara Howes. Finally, I would like to thank my wife Sara and daughter Rhianna for their continued support.

Contents

Chapter 1
Forecasting Methods: the Case of Horseracing

Forecasting the outcome of given situations provides a means of satisfying our natural desire to know the future. In almost every commercial environment there is a need to provide forecasts, from likely crop yields in agriculture to the value of Sterling in the field of finance. The demand for more accurate predictions, coupled with the advances in computer technology, has led to dramatic changes in forecasting methods. It is now possible to supplement traditional statistical forecasting techniques, which have been used for many decades, with rule-based and knowledge-based approaches, which utilise the knowledge gained from human experts in their construction, and machine learning methods in which the computer *learns* from available examples without significant human intervention. This book examines the application of these forecasting methods to the domain of horseracing in Great Britain with the aim of producing a range of techniques which can be used to forecast the results of horse races.

Forecasting

Forecasting concerns the relating of an outcome to a specific set of circumstances, a 'conjectural estimate of something future' according to the *Concise Oxford Dictionary*. Humans make many forecasts, or predictions, every day, whether in regard to the likely position of an oncoming vehicle, or the possible result of a sporting event. However, whilst the latter would be unanimously accepted as a forecast, the former would not necessarily be viewed as such. This illustrates two facts: the high degree of diversity associated with forecasting techniques and the importance of this deductive process. The diverse nature of forecasting results from the many different approaches employed in different situations. For example, the fact that the sun always rises in the east, and can be predicted as doing so, can be explained by the laws of planetary motion. However, a knowledge of this theory is not the only basis on which the prediction could be made. For instance, someone who had seen the sun rise in the east for the previous 25 years might also conclude, with as much certainty as the physicist, that the sun will rise in the east tomorrow. These two approaches could hardly be more diverse, however both are valid methods and generate equally accurate predictions. The fact that we devote so much time to forecasting is evidence of its importance, and, furthermore, the way we make our predictions has a bearing on our view of events. Casti[1] asserts in the book *Searching For Certainty* that 'making sense of the things we see and predicting the future course of events have always played an essential role in the formation of each individual's world view'. In essence, forecasting is part of our lives and our ability to make accurate predictions is essential to our survival.

Sporting events, especially horse and greyhound racing, are designed to encourage the public to form an opinion about the event and to express this opinion in the form of a bet. After all without betting neither of these two sports would exist. However, unlike the National Lottery and other completely random numbers games, where attempting to forecast the outcome is a pointless task, horseracing poses the race analyst with a challenge that can vary in terms of complexity from the (apparently) trivial to a level commensurate with the most testing *Times* crossword.

It is interesting to note that this level of complexity, although a product of the task itself, is also linked to the analyst's knowledge of racing. To the novice racegoer selecting the probable winner from a field of 24 runners is simply a matter of identifying a well-known jockey or appealing name. To the experienced race bettor such a race may take hours of intensive form study before a conclusion is reached, unfortunately, with no guarantee that this painstaking work will yield a more accurate prediction than the former approach for a single race.

Whichever approach you choose, the horseracing problem offers a challenge worthy of our best efforts.

The Horseracing Problem

Thoroughbred horseracing in Great Britain has been extremely well documented for many years. Records still exist detailing the very first organised horseraces, such as the Newmarket Town Plate in which King Charles II was successful in 1671. Consequently, a wealth of information is available to the race analyst, concerning all races and the horses that compete. The historical records include the peculiarities of each race track, whether right or left handed, undulating or flat. For each horse, all previous race details including times, race distances, course conditions and race commentary, together with the animal's pedigree are documented. Jockey and trainer statistics are also available indicating success rates by several variables including track and race type. The fact that racing is so well documented is helpful to the race analyst by providing the basic information with which to work.

Although the availability of this large volume of data is advantageous for modelling purposes, the level of information relating to each horse in a race is extremely detailed which severely complicates the task of generating workable systems. Discussion of race analysis methods with recognised racing experts suggests that this level of detail, coupled with a lack of structured approaches to race analysis, has resulted in wide disagreement between the experts regarding optimal solution methods. Although there is general agreement between the experts in the identification of horses with either very high, or very low, probabilities of success, there is considerable disagreement for less well-defined runners. Interestingly, these problems mirror those found when developing computerised methods for other prediction problems, such as assessing mortgage applications[2].

A second problem with the horseracing domain is the competition between the runners in a single race. A horse may possess outstanding winning credentials, but the likelihood of its success is also dependent on the abilities of the other runners in the race. This comparison between the animals is a problem even the experts find difficult to handle. For example, a well-known race analyst discussing selection techniques commented: "and now the guessing starts" when faced with comparing several animals with similar credentials. This competitive element must be considered in any horserace forecasting model in both the output of the system and in the data used to construct the model.

In addition to the vast number of example cases, a major characteristic of the data is the high level of uncertainty associated with many of the components used in the modelling process. This uncertainty is due to the methods used to determine the attributes of the animals on which the forecasts are made. Whilst accuracy is possible with respect to variables such as age, others are subject to measurement error, and some rely totally on opinion. This, naturally, increases the complexity of any modelling procedure. Furthermore, in some cases data will be missing, and to make matters more complicated, a high degree of inter-correlation can exist between the variables. For instance the best jockeys tend to ride for the most successful stables and owners, an important consideration which should certainly not be overlooked, especially when constructing betting systems.

So we are faced with a complex problem, which although well-documented, comprises missing and uncertain data, complex inter-relationships and an element of within-race competition. But it is not an insurmountable problem, and a range of techniques exist to provide a solution.

Intelligent Systems

Many computer systems have been developed that are labelled intelligent. For instance, chess programs are now capable of beating even the very best players. However, any assessment of machine intelligence is dependent upon the definition of the word *intelligent*. Sharkey and Brown[3] argue that, in the main, programmed solutions simply reflect, but do not possess, the intelligence of a human. Since the programs do not produce the solution method themselves they cannot be thought of as intelligent. Other definitions simply require the system to exhibit a level of *understanding* to be labelled intelligent. Understanding implies a depth of knowledge about a specific issue from which, given a reasoning strategy, a conclusion or explanation may be derived. This is where conventional programs and systems categorised as *artificial intelligence* differ. Traditional systems do not exhibit any apparent understanding of the problem and are generally algorithmic: each instruction is performed in an order determined by the program code. In contrast, some artificial intelligence systems are not constrained by rigid algorithms that dictate the order in which the instructions are performed, and they are also able to demonstrate useful reasoning strategies, thereby exhibiting an apparent knowledge of the domain.

Three different forecasting approaches using the ideas and theories of artificial intelligence are examined in this book: *rule-based methods, knowledge-based systems* and *connectionist approaches*. Of these methods, the first two, rule-based and knowledge-based techniques are already widely used by race analysts. However, connectionist systems have so far been all but ignored.

As the name suggests, rule-based methods rely on the formulation of rules on which the prediction is based. In horseracing this type of selection method is also referred to as a *system*. Systems provide the bettor with a rigid set of rules to apply to each race in order to determine whether a bet should, or should not, be made. The rules normally take the following form:

if *Condition A is true*
and *Condition B is true*
then *bet*

In this example, the bettor needs to determine whether the two conditions, A and B, are met, if so the bet should be placed. Complex systems have large rules with many antecedents which need to be satisfied such as in the following rule which is designed for American flat racing[4]: