

Perl による競馬予想支援システムの構築

(環境計測) 堀井智行

1 はじめに

近年、ディープインパクト号の活躍により、競馬人気が高まっている。予想的中する事は、競馬を楽しむ上で大きな要素である。本研究では、競馬の予想をする上で有用なデータを web 上から自動で収集し、オリジナルデータベースを作成し、レースの結果予想を支援するシステムの構築を目指した。

競馬の予想において考慮するファクターは様々であるが、最も端的に競争馬の能力を推し量る指標として、走破時計(スタートからゴールするまでのタイム)の比較が挙げられる。しかし、走破時計は、レースの展開、馬場(コース)状態等に左右されるので、単純に走破時計を比較しても、それは競走馬の能力比較にはならない。

中央競馬会(以下 JRA)主催の競争は、芝コースとダート(砂)コースを使用して行なわれる。芝コースは、芝の生育状況の季節変化等により、コース状態を厳密に把握することは困難である。ダートコースは、同じ競馬場に限りみると、コース状態を決定する主要なファクターは水分量であると言える。ダートコースの馬場の状態は JRA により水分の含有量の違いによって、良(水分 10%未満)、^{ややおも}稍重(10%~13%)、重(13%~16%)、不良(16%以上)の4段階で発表されている。ダートコースの場合、馬場の水分量が多い時の方が走破時計が早くなる傾向があると一般的には考えられており、このことは、ダートコースにおけるコースレコードタイムの殆どは馬場が水分を多く含んだ時に記録されていることから確認できる。しかし、主催者発表の4段階では、「良」時の水分含有量の区分が「稍重」「重」の3%刻みに比べると不十分であるように思える。そこで、この「良」馬場時の走破タイムを前日に降水がある場合、当日に降水のある場合、2日間全く降水の無い場合に分けて分析した。その結果を図に示す。この図は、2006年1年間に京都競馬場ダートコース1,200mで行われた獲得賞金1,000万円以下の馬による競走についての、レースごとの全出走馬の平均タイムである。レース条件の選択理由として、1,200mは、最も短距離であり、レース展開によるタイムへの影響が出にくいと考えられることから、また、獲得賞金1,000万円以下の馬によるレースは、出走頭数が他の条件より安定していることから選択した。この図からもわかるように、JRA 発表では「良」とされる馬場でも、降水の

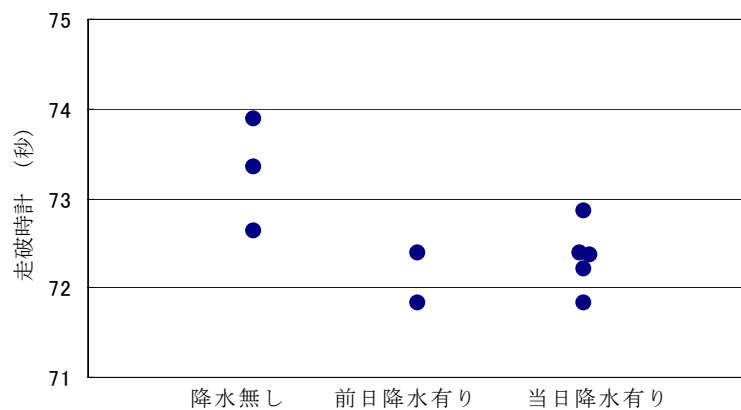


図 「良」時の走破時計の降水量による分布

ある場合は走破時計がより速い方向に推移している。このことから、馬場状態を考慮した上で走破時計を比較するには、少なくとも、水分を含んだ「良」か、そうでない「良」の区別が必要ではないかと考えられる。よって、主催者発表の4段階ではなく、レース前からの降水量を記録することで、馬場状態をより厳密に把握し、降水状況と競走成績と結合したデータを蓄積、表示することにより、走破時計による競争馬の能力評価を支援するシステムを構築した。

2 システムの概要

今回作成したシステムは、毎週行なわれる競走に関して、レース2日前からの降水状況と競走成績を結合したテキスト形式のデータを蓄積し、そのデータベースを基に、ブラウザからの検索条件入力に対し、CGIを起動し、過去の競走成績を出力する。蓄積するデータは、今後の機能拡張性も考慮し、馬名、レースの日時、競馬場名、レース番号、距離、獲得賞金別クラス、走破時計、ゴール前3ハロン(約600m)の走破タイム、負担重量、降水量(2日前からの毎時降水量)の10種類とした。競走成績及び降水状況は、既存のwebページに公開されている情報を自動で取得する方法を採った。また、全国の競馬場で行なわれる全ての競走について網羅したいが、今回は京都競馬場及び中山競馬場に対応したシステムを作成した。

サーバを構築するマシンは、CentOS4をOSとし、webサーバソフトウェアにはapacheバージョン2.0.52を、プログラム言語には、perlを使用した。perlは、CGIやテキスト管理、システム管理に広く用いられているライセンスフリーの言語で、特にテキスト処理能力に優れている。また、様々なライブラリ、モジュールが確立されており、複雑な処理もライブラリやモジュールを利用する事で簡単に実行できる。今回はCentOS4パッケージに標準装備されているperlバージョン5.8.5を使用した。

2-1 データベース作成スクリプト

Perlに搭載されているモジュールを利用し、レースの当日及び2日前(木、金、土、日)の毎時降水量を気象庁ホームページより取得した。中山競馬場の降水量として、舟橋市のデータを、京都競馬場の降水量として、長岡京市のデータを採用した。HTMLソースファイルから、必要な文字列のみを抽出、整形後、降水量データベースとしてテキストファイルに追記していく。同じ方法で、yahooが提供する競馬情報ページより競走成績を取得し、必要な情報を抽出、整形する。ここで、降水データベースより降水量の毎時値を読み込み、競走成績の後ろに羅列する。これをテキスト形式で競走成績データベースとして蓄積する。

2-2 入出力スクリプト(CGIスクリプト)

馬名及びレース条件の入力画面(HTML)への入力に対し、競走成績データベースを読み出し、入力された馬名にマッチングした行から、日付、レース番号、着順、走破時計、降水状況をブラウザに出力する。また、馬名に対する出力を走破時計順にソートするボタンを搭載した。また、レース前2日間に降水が無い場合、馬場状態が同一であるとみなし、前2日間の降水量が0の場合のみの成績を表示させるボタンを搭載した。