

ボートレース結果のデータベース化と解析

理工学部数理情報学科
T980121 守屋 洋明
指導教員 佐野 彰

概要

私たちの身の回りにはたくさんのデータベースがある。電話帳でも、新幹線の座席予約システムもひとつのデータベースである。データベースはデータの集まりから構成されている。データベースの利点は大量のデータから追加、削除、更新、検索といった操作が素早く正確に行えることである。データベースは扱うデータの量が多ければ多いほどますます威力を発揮できるのである。

本研究では、ボートレース結果をデータベース化して「進入コース別の着順」、「選手がどの進入コースをよく選ぶか」、「選手のスタートタイミングの平均値」を調べ、解析した。

まず、ボートレースのレース結果を「競艇オフィシャルWEB」からダウンロードした。次にダウンロードしたデータファイルをPerlによって取り出しSQL用に数値化してデータベース化した。更にJavaによるSQLインターフェイスを作成し、データベースの中の200万行近いデータの中からレース結果を自由に取り出せるようにした。

以上の環境を構築した上で、そのデータベースの中から「進入コース別の着順」、「選手がどの進入コースをよく選ぶか」、「選手のスタートタイミングの平均値」を調べた。その結果、開催場によって進入コース別の着順が大きく違っていることがわかった。1コースから1着率は全国平均では29%だが大村では36%もあり、3回に1回以上1コースからの進入で勝つことがわかった。さらに2着以内である連帯率では1コースからの進入で6割となっており、1コースから進入する選手に賭ければ資産を増やすことも可能であるということがわかった。それに対して琵琶湖では1コースからの1着率は23%、連帯率は46%しかなく1コースからは非常に勝ちにくいことがわかった。

開催場別にそれぞれの進入コース別の1着率や連帯率、選手がどの進入コースをよく選ぶか、その進入コースにおける平均スタートタイミングを調べた。その結果、例えば1コースからのスタートタイミングが早い植木通彦選手が大村で1コースから進入すれば3回に1回以上に1着になることがわかり、6割以上の確率で2着以内に入ることがわかった。

また、相関係数を調べた結果、着順と一番相関が強いのは進入コースであるということがわかった。着順とスタートタイミングではそれほど相関は強くなかった。自然条件である風速や波の高さは着順に大きく関係あると思われたが、相関係数を調べた結果、全くと言っていいほど相関は見られなかった。よって、着順を想定する上で一番参考になるであろうデータは進入コースであることが明らかになった。