

研究テーマ: データマイニングを用いた野球選手の成績分析

名列番号 024 氏名 新平 智紀

1. まえがき

データベースなどに蓄積されたデータを、マイニング(採掘)し、宝物である情報・仮説知見・課題等を見つけ出すことをいう。分析結果をもとに法則性の発見、効果測定、予測など様々な要因を導き出すことが可能。結果をもとに様々な対策が戦略的にとれる。

2. 研究課題

野球選手のシーズンの成績をデータベースとし、そのデータを分析し、新たな知識や法則性の発見を行ってみる。

3. 研究方法

野球選手の成績のデータを集める

選手の成績ファイルを作成するプログラムを考える

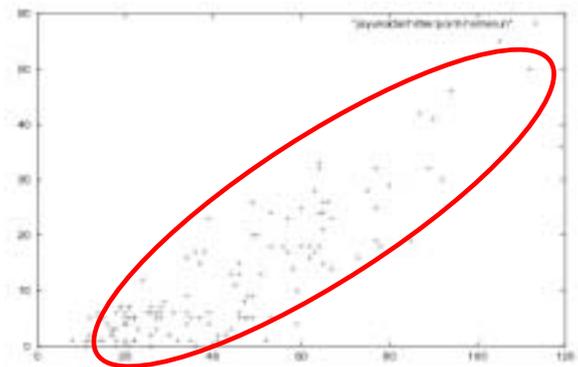
選手の成績を入力しファイルを作成する

成績ファイルをグラフにして視覚化する

グラフから新たな知識や法則性の発見を行う

4. 実験と考察

得点と本塁打のグラフを作ると



のようになります。これより、得点の多い人は本塁打も多いということが分かります。

これらのいろいろな要素について調べると、次のようなことが発見できました。

- 得点(ホームベースを踏んだ回数)が多い打者は、三振の数も多い傾向がある。
- 得点が多い打者は、本塁打と塁打が多いという傾向がある。
- 本塁打が多い打者は三振も多い傾向がある。
- 盗塁が多い打者は併殺打が少ないという傾向がある。

5. まとめと今後の課題

- データマイニングに必要なもの:
「データ収集力」、「データ処理力」、「データ蓄積力」、「データ発掘力」、あとは「体力」。
- データの取舍選択、集計、成形など地道な作業を繰り返し行わなければいけない。

データマイニングにおける失敗の原因

- 無秩序なデータ収集:
「何か重大な情報があるかもしれない」とやみくもにデータ収集を行うと、解析レベルの低下のみならず、逆に身動きがとれなくなる。
- データマイニングを魔法の技術と勘違いする:
データマイニングを用いれば何でも情報が発掘できると勘違いしてしまうこと。

最後のまとめとして、

データマイニングにおいて、コンピュータは、人間と違って、疲れがない、処理速度が速い、計算間違いをしない、など必要不可欠なものであるが、あくまで目標となる指標をだすための補佐的な役割であって、最終的に意思決定するのは、人間である。

参考文献

- (1) 山本 喜則/松井 洋子 著, 情報処理のためのCプログラミング, 森北出版, 1996
- (2) 椋田 實, はじめてのC, 技術評論社, 1993
- (3) <http://sports.yahoo.co.jp/baseball/teams/>
- (4) <http://www5.ocn.ne.jp/~shinya91/>
- (5) <http://www.icot.or.jp/ARCHIVE/Museum/SOFTWARE/KAPPA/pdst/node4.html>