Owen FEIK

Bowler Selection Data Model

Table of Contents

[Introduction 1](#_Toc2539389)

[Problem 2](#_Toc2539390)

[Assumptions 2](#_Toc2539391)

[Inclusions 2](#_Toc2539392)

[Exclusions 2](#_Toc2539393)

[Data Modelling 3](#_Toc2539394)

[Entities 3](#_Toc2539395)

[Model 4](#_Toc2539396)

# Introduction

The algorithm for which this data model is designed is intended to determine the ideal bowler to bowl out a batsman in a game of cricket, given a number of statistics about the bowler, batter and superset of each of those types.

This data model is designed around the idea that the optimal solution for recommending bowlers focusses not on the raw statistics of the players, but on trying to quantify players style and comparing accordingly. A player’s style refers to the particular, unique, technique that player applies to their game.

It uses a combination of structures to allow data to be quickly accessed and manipulated, as well as practically stored.

To best do this, it was decided to have two main entities, the bowler and the batsman. Each of these would have a number of pieces of information relevant to their style associated with it.

# Problem

## Assumptions

This solution assumes that the optimal way to approach this problem is through the players style. In doing this, it ignores the possibility of comparing players along raw statistical values such as batting average. This is done because it is felt that these statistics are more representative of the players individual skill than of the particular characteristics that make one bowler superior to another in a given situation.

## Inclusions

The model developed makes use of a number of the statistics listed on the HowStat website. These values have been chosen because it is felt that they best represent the style of the player as opposed to (or in addition to) their skill level.

For batsmen:

* Hand
  + The hand a player uses can be relevant to which bowlers most affect them.
* Aggregate
  + This value, in conjunction with the number of 4s and 6s is used to determine the batsman’s rate of 4s and 6s, determining their tendency to hit long range shots.
* 4s
* 6s
* Scoring Rate
  + This value can indicate how conservative/aggressive the player is, and has, therefore been chosen as a tertiary indicator of style.

For bowlers:

* Hand
  + Left- or right- handedness has an affect on the style of the bowler.
* Style
  + A bowler’s style is defined as either spin or pace, each of which are substantially different styles, which a batsman might fare better or worse against.

## Exclusions

Other statistics have been excluded because it is felt that the values listed above are those that best represent the style of a given player.

While statistics like batting average are extremely relevant to the quality of a player, it was felt that they were less relevant to the *style* of play, which was seen to be the most relevant aspect for the suggestion of a bowler.

# Data Modelling

## Entities

Batsman

An entity which describes the batting style of a player.

Attributes:

* Name
  + The full name of the player, for identification purposes.
* Hand
  + Used as the primary indicator for style comparison between players.
  + The hand the player uses for batting.
  + Right, Left or Ambidextrous, which can be represented as 0, 1 or 2.
* Percentage of hits that are 4s/6s
  + Used as secondary indicator for the batter’s style.
  + Number from 0-100.
* Scoring Rate
  + Tertiary statistic for comparison.
  + Number.
* Conceded wickets
  + Used to determine which bowling styles have been most effective against the batsman.
  + The list of bowlers that have bowled out this player.
  + List of names.

Bowler

Entity which describes the bowling style of player.

Attributes:

* Name
  + The full name of the player, for identification purposes.
* Hand
  + First part of style indication.
  + The hand with which the player bowls.
  + Right, Left or Ambidextrous, which can be represented as 0, 1 or 2.
* Spin/Pace
  + Second part of style indication.
  + Whether the player primarily focusses on spin or speed of the bowl.
  + Spin or Pace, which can be represented as 0 or 1.

## Model

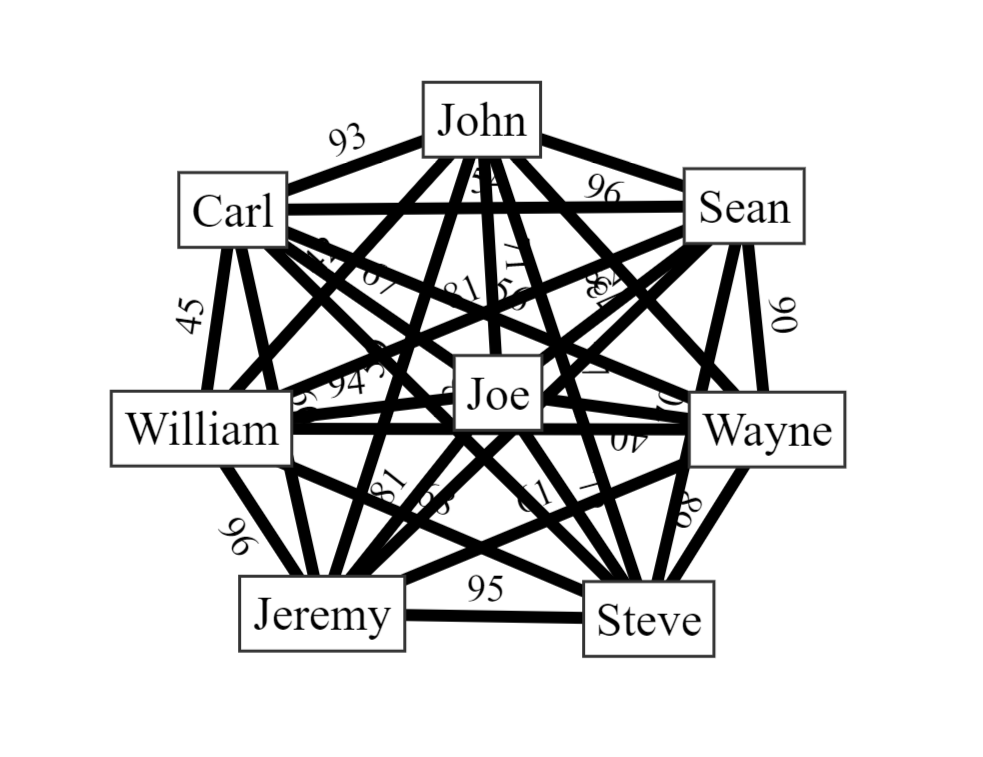
The problem is modelled as follows:

There exists a set of all bowlers and a set of all batters – a player can be in both categories.

* + - These two sets can be implemented as lists, a list of bowlers and a list of batters.

For each batter in the set of batters, a number is generated that represents the degree of similarity between the style of the batter and each other batter.

* + - This can be represented as a graph where each node is a batter, and each batter has an edge to every other batter. These edges are labelled with the similarity figure of the two batsmen.



Then, to select a bowler for a given batsman, we locate the batsman on the graph and apply an algorithm that considers the performance of that batsman against right/left handed and spin/pace bowlers, as well as the performance of each other batsman against them, using the similarity value as a coefficient to increase the influence of more similar batsmen.

At this stage we are left with a score for left/right handedness and spin/pace bowling, which will indicate which of these is preferred.

Finally, we select the most appropriate bowler from the subset of bowlers who are available by choosing the one whose style most closely resembles the result of the previous algorithm.