

Team Project - An Introduction Algorithms for Sports Eliminations

Gordon Reid - 1002536r

February 26, 2013

Preliminaries

The project is based around working out if a baseball team has been eliminated from the division it is currently playing in however no knowledge of baseball is required. The project can also be extended to some other sports such as Basketball. Each division has a small number of teams who play each other multiple times. Each match requires one team to win. The winning team earns a single point, whilst the losing team does not earn any points. Draws are not permitted. A team is eliminated if it cannot earn enough points to secure the position at the top of the points table. For the project, American leagues and divisions will be focussed on as baseball is popular in America.

The report makes little use of jargon terms, and has been structured with the aim of being as accessible as possible. Parts of the report make reference to terms and concepts where knowledge of Software Engineering, Object Oriented Programming, Graph Theory, and Network Flow will be useful but not required.

For those readers who may be unfamiliar with Graph Theory, I will briefly describe the basic structure of graphs, in particular directed graphs. Graphs are made up of two components, vertices and edges. Vertices can be thought of as cities, with edges being roads between the cities. A road between two cities can have a maximum number of cars on it at one time (its capacity) and a number of cars on it at the present time (its flow). The flow of traffic can never exceed capacity. The road from city A to city B is separate from the road from city B to city A and the existence of one road does not imply the existence of the other.

Network Flow is the study of working out how to have as much traffic move from a start city to an end city by passing through intermediate cities.

There is only one algorithm discussed in the report, this is the Ford-Fulkerson algorithm. The algorithm involves creating graphs for each team from the remaining matches between two other teams in the division and the other teams themselves. The team being tested only has a mathematically possible chance of winning the league if there is the flow from the special vertex 'source' to the other special vertex 'sink' is saturating. Saturating meaning that the total flow out from the 'source' is the same as the total capacity out from the 'source'. If there is no saturating flow then the team has no chance

of winning. The source vertex is a vertex where there are no incoming edges (start city), and the sink vertex is a vertex where there are no outgoing edges (end city).

Outline

The remainder of the report will cover the background of the project and will explore the reason why the sports elimination calculation became something of interest. The background section will also give two references to publications the project is based and builds on.

After the background material, the report will then discuss the aims of the project, from the primary goal to the extensions that will be explored if time permits.

Finally, our motivation for working on this project will be discussed.

Background

The project builds on two papers=: one by K. D. Wayne, ‘A new property and a faster algorithm for baseball elimination’, SIAM Journal on Discrete Mathematics <http://epubs.siam.org/doi/abs/10.1137/S0895480198348847> and one by I. Adler, A. L. Erera, D.S. Hochbaum and E. V. Olinick, ‘Baseball, optimization and the world web’, manuscript, 1998 <http://riot.ieor.berkeley.edu/~dorit/pub/baseball.ps>.

The project came about from the American love for baseball, and the love for a multitude of statistics relating to baseball. A common news story for sports fans is the announcement that a team can no longer win the current season however the calculations used to work this out are typically naïve and do not take into account the remaining schedule of games.

The papers stated above, build on the naïve calculations and make use of a simple, yet effective, algorithm to reliably answer the question: ‘Is it mathematically impossible for a given team to win the baseball league?’ The algorithm is known as the Ford-Fulkerson algorithm, named after L. R. Ford, Jr. and D. R. Fulkerson. The Ford-Fulkerson algorithm removes the need for an exhaustive search over all of the possibilities of results the remaining matches can have.

Aims

The primary aim of the project is to implement the aforementioned Ford-Fulkerson algorithm that, given a league table, remaining schedule of matches, and a team, will determine whether or not the given team has a chance of finishing at the top of the league table. The results will then be displayed on a user interface with data being obtained from either a text file or parsing dynamic up-to-date information from a sports website. The user interface will be fairly simplistic with a large central table displaying the league and elimination information with buttons and/or radio buttons for navigating between leagues. The user interface may also allow the user to navigate between weeks

to see how the league progressed and when teams were first eliminated. A web-based version of the user interface is of interest and will be explored.

There may be various ways the given team can be eliminated from the league and a secondary objective is to provide a ‘certificate’ of elimination. This will provide the team, or teams, that are responsible for ensuring the given team cannot finish top of the league table. A simple example can be taken from the naïve calculations traditionally used by sports pundits. The teams responsible for the elimination is just the list of teams that have a points differential with the given team that is greater than the number of games the given team has to play.

At time of writing, the algorithm has been completed and has been extended to show the teams responsible for eliminating a given team. The user interface is also largely complete as all currently available information is being displayed to the user. The previous/next week buttons currently offer no functionality as this is still to be implemented. The simple static text file parser is complete.

The next steps are to create a web-based interface that will pull data from a MySQL server’s database on page load. This is currently at the functional prototype stage with a sample web page showing a simple textual output of all division league data but will soon be extended to four pages showing the data in a tabular format.

As an aside to the project, further statistical analysis will be discussed. With baseball being one of the most statistically analysed sports it may be possible to implement a certain confidence level indicator for teams which still have a mathematically viable way of winning the league. If a given team is known to win against the teams in the matches it has remaining then it can be deemed more likely that the team can win the league, and vice versa. This is not possible with the Ford-Fulkerson algorithm and is very much an interest rather than a requirement.

Motivation

The project is very interesting to me. I may have no knowledge or interest in baseball itself however I do enjoy seeing algorithms and theories used in a real world scenario.

The results that can be generated from the project are also interesting to baseball fans who may want to know how well their team is doing in the current league, and to see if any major competitors have been knocked out of the running. This will become increasingly relevant towards the end of the league as a clear winner, or possible winners, can become apparent.