**CS408 Individual Project**

**Craig Thomson**

**11/2/15**

**Project Progress Report Submission**

**Conn03 – Who will win the tennis/football/cricket?**

**Revised initial goal**

The initial goal of the project is to predict major game events from sports games, from betting market data, and relay it’s predicted information to the user, for example it try and relate large changes in betting odds to events in game. This aim of it is to provide a free live match feed giving out key information to observers, whereas alternative programs providing the service are typically not free. The other aim is to discover how much information about an in-play sports game can actually be extracted from the betting market data. The program will run whilst a game is playing (the user selects what they want to look at) and it will attempt to predict events in the game in near real time (5-30 second delay is the plan). The program will primarily focus on football but may be able to support other sports.

**Revised achievable goal**

Within the time frame it’s definitely possible for the program to work well with football. The aim is to add support for another sport, its analysis won’t be as well developed as that, that will run for football but it should at least be able to tell users who wins each set.

**Alterations from previous expectations**

A big change to the project is that it will now support 1 or 2 sports very well (well in the sense that more events can be extracted) rather than poorly supporting multiple sports. Some sports at first glance, and after referencing related work look impossible to model well (horseracing), others look unfeasible (golf) and others look promising but I haven’t looked into them yet (darts, tennis and snooker). American Football was initially a sport that I wanted to model but that’s impossible now since the season has already ended and it would be challenging to test, to say the least.

**Progress achieved**

So far the project is at a point where data for football is being collected. Currently there’s a class in the project that polls the BetFair API for data for a selected game and outputs timestamps with its predicted probability (from the hypothesis that probability is the number between the highest available back price and the lowest available lay price at a point in time). Data is currently being collected, after approx. 10 games of data has been collected I will graph it and compare the changes in a team’s probability (initially looking at the market regarding who wins the game) to a websites list of time stamped match events to look for relationships between the changes in probability and events occurring.

**Revised Project plan**

**12/2/15:** Finish prototype that collects game data and stores individual runner data in csv files

**13/2/15:** Project Progress Report

**15/2/15:** Finish collecting football game data

**15/2/15:** Finalize GUI design mock ups

**17/2/15:** Investigate relation between odds changing and game events

**21/2/15:** Finish implementation for modelling Football

**26/2/15:** Look into other game markets to extract more additional information for football

**28/2/15:** Finish final football implementation

**1/3/15:** Investigate GUI creation choices

**5/3/15:** Complete initial implementation of GUI

**6/3/15:** Investigate other sport modelling

**10/3/15:** Complete other sport modelling

<Optional time here to either further the modelling or add additional sport support>

**13/3/15:** Project Report outline and Draft Chapters

**7/4/15:** Project Report Due

**20/4/15:** Bound Copies of Project Report due

**Introduction**

In recent time there has been a large increase in the popularity and revenue generated from sports gambling. There is a large number of online based betting companies that exist, and will be created in the future, to facilitate user’s desires to bet on a large range of markets and to get the opportunity to bet on new games. Online betting markets have expanded over time, initially starting with universally popular sports in certain regions, such as football, tennis and cricket. Over time the number of available betting markets has increased, now with major betting websites providing support for many sports to be betted on, and new websites being created to support betting on new niche markets.

Along with the increase in betting popularity, it means that the betting markets provided have become more popular, with large amounts of money being transferred through the duration of the game. With these more dynamic changing markets, prices are changing constantly through the duration of the game.

The aim of this project is to use the dynamic near real-time market odds extracted from a betting website to see if its information-equivalent to what’s occurring in the game, over time as the odds and game state change. It does this by feeding itself data at certain intervals from a popular betting websites API. Doing this successfully will also provide users a free program that provides users a near live feed of game events. Other services that typically do this are for a price.

Chapter 2 of this report provides an in-depth look into material related to the project aim. Chapter 3 goes on to give a formal presentation of the problem description and specification.

**Related Work**

This chapter of the report looks into work related to the problem of trying to model a sports game as a set of probabilities and trying to extract meaningful information from it. There has been a large amount of work that involves using BetFair and collecting game data for analysis, but most of this work has been used to attempt to predict market trends and outcomes in order to find points in a game where it’s most financially beneficial to put bets down, rather than trying to model the games as probabilities and try to predict game events.

**Using BetFair for modelling Tennis from probabilities**

There is a similar performed by a student from University College London. They used the BetFair API to initially collect tennis game data and create a model of the game. They then created a program that would collect live data and infer the score line from the data. Because of the nature of the tennis their goal was only to predict who wins sets and in the end, the game, from the shifting odds.

**BetFair API**

BetFair provides a powerful free API that has multiple purposes. It was released so that app developers could create their own applications which can be used to allow users to perform any actions that can be performed on their website, so that they can keep users and possibility bring in new ones if they disliked the original UI given by BetFair. The API supports all website operations and the access of live match data. The API has a fairly large data limit so that lots of data, possibly regarding multiple markets/games with many filters can be collected in a single request. BetFair also provides a lot of mostly correct documentation that introduces users to it and they have a GitHub page with examples of their API in action using a diverse range of programming languages.

**Problem Description and Specification**

The goal of this project is to see exactly what game events can be inferred from live shifting probabilities and how much detail can be extracted. For example with the game of football, the goal is to see what events can be extracted and how precisely they can be extracted, in terms of the extracted time to their actual time (if it correctly predicts the event). A basic solution that works with football would extract when goals occur in a game, which team scored and a fairly accurate timestamp. A more complicated solution would be able to extract events like when cards are issued, corner kicks, free kicks, exactly how long the two game halves took and penalties issued.

The current project spec aims to be able to extract a subset of major game events. The plan for it is to be able to predict cards, when halves end, goals and penalties. In order to create a solution the first step is to collect data from the key game market (Match Odds). From there the shifts in odds in the Match Odds market will be collected, graphed and the changes in odds compared to a list of game events. Events in games such as penalties, cards and when halves end will need to analyse additional markets for the information. Since obviously if a penalty is awarded and if it’s successful then it is noticeable in the Match Odds market, but if unsuccessful it may have no effect in odds, if the market is sampled at long intervals.

The plan for the project is to collect a range of game data for different markets, identify reasons for odds shifts and then write a program that will realise these specific changes in real time and make predictions from it.

The project aims towards accurately predicting events for large, popular betting games, rather than smaller games due to the number of markets available, and thus possible amount of data that can be extracted.

**Current State**

Currently the project is just a program allows the user to pick a sport, then a game and then a market that they want to track. It then waits until the game starts and then polls its market data every 30 seconds until the game stops and then it saves the runners data for the market into a csv file. Currently the csv file is graphed and then it’s manually compared to a list of the game events from an external website to match up changes in odds to actual game events.

**Verification & Validation Strategy**

There’s only really one way that the program can be tested, which is comparing its output to what occurred in the game. There’s 2 ways to do this, firstly it would be to watch the game and compare what happened at a certain time to what the program said happened at that time. This test strategy is quite time consuming so the other alternative will be used. The other alternative is to get a list of game events that occurred from a trusted website and compare what it says happened at certain times to see what the program produced. The programs GUI won’t need much testing when it comes to usability. There’s only approximately 3 screens the user needs to go through to go through until you get a screen that gives game results. The game result screen will be based off the standard way of displaying a football game, as used in many football computer games and online game previews, with the top half of the screen separated into 2 parts, one for the first one and the other part for the second. There will be a box displaying game events with information regarding time, event and what team performed it.