Price Prediction in Sports Betting Markets*

Juan M. Alberola**, Ana Garcia-Fornes, and Agustin Espinosa

Departament de Sistemes Informàtics i Computació,
Universitat Politècnica de València, Camí de Vera s/n. 46022, València, Spain
{jalberola,agarcia,aespinos}@dsic.upv.es

Abstract. The sports betting market has emerged as one of the most lucrative markets in recent years. In this kind of prediction market, participants trade assets related to sports events according to their expectations. Prices in sports betting markets continually change depending on what is happening in the event. In this paper we propose an approach focused on predicting price movements in order to make benefits regardless of the final result.

We develop an agent who participates in the market focused on the task of learning the price movements in order to make predictions of future prices. Our approach is based on identifying and learn pattern price movements in order to predict the price movements of new events by using an underlying Case Based Reasoning system.

1 Introduction

Prediction markets are speculative scenarios where participants make predictions about future events. Assets regarding all of the possible outcomes of the event are created, and the price of these assets is related with the probability of each outcome. Participants exchange these assets according to their expectations with other participants or with a bookmaker.

In the last few years, sport betting markets have emerged as one of the scenarios in which the most money is exchanged everyday. Sport betting markets are a specific kind of prediction markets where the traded assets are referred to sporting events. Therefore, the attraction of betting on sporting events and the growth in popularity it has experienced, has meant that millions of users make more exchanges in sports betting markets in an average day than other exchanges scenarios such as other financial markets[1].

Sports betting markets usually have a short or very-short duration in comparison with other prediction markets, such as political markets. Markets regarding the probability of landing on Mars in ten years or the probability of a particular candidate becoming the next US president can last months or even years. However, markets regarding the winner of a soccer game or a horse race usually

* This work has been partially supported by CONSOLIDER-INGENIO 2010 under grant CSD2007-00022, and projects TIN2008-04446 and PROMETEO/2008/051.

** Juan M. Alberola has received a grant from Ministerio de Ciencia e Innovación de España (AP2007-00289).

J. Dix and C. Witteveen (Eds.): MATES 2010, LNAI 6251, pp. 197–208, 2010.

last a few hours or minutes. Therefore, prices usually tend to change quickly according to every influential factor related to the event.

Studying how probabilities, and therefore prices, change during the sporting event will allow us to approach sports betting markets in a new fashion more similar to the financial markets. As financial traders buy and sell assets according the prices and their expectations of price increases or decreases, assets regarding sporting events can be traded at a given price in order to make an opposite trade later at a better price, with the goal of making profits regardless of what the final outcome is. Thus, in these markets it is important to predict future price in order to identify the best trading decisions.

Price movements are made by the participants of the sports betting market according the probabilities of the outcomes change during the event. Therefore, participants change prices in a specific sporting event according to what happens in the event.

Based on this idea, we are interested in identifying price movement patterns that can be repeated in different events under the same underlying circumstances. To achieve this goal we develop an agent with an underlying Case-Based-Reasoning (CBR) [8] in which by means of observing past sporting events, the agent is able to predict future price movements for an unknown sporting event and therefore, drawing a price evolution over time.

The rest of the article is organized as follows. Section 2 presents previous works related to our proposal. Section 3 describes the Sports Betting Markets and the trading possibilities according to price movements in order to make profit. Section 4 details the market structure. Section 5 describes the structure of the CBR-agent. In section 6 we evaluate the prediction accuracy of the CBR-agent with real data. Finally, section 7 discusses the contributions of the paper and future work.

2 Related Work

Prediction markets have been studied as powerful mechanisms for predicting the probabilities of future events [5]. Most of the research on prediction markets is focussed on pricing, that is, assessing the most accurate price according the probability of the event. Studies regarding prediction markets, such as the Iowa Electronic Markets, the Foresight Exchange or the Hollywood Stock Exchange demonstrate than these markets provide very accurate probability predictions of future outcomes [11,2]. Other works are focused on stuyding how information is incorporated into the market and therefore, influences the prices [3]. However, to our knowledge little effort has been made towards studying price evolution in prediction markets, with the aim of making a profit regardless the correct outcome.

The problem of predicting future prices and price movements has been broadly studied in the economic field. Since the introduction of computational tools for modeling financial and economic markets, several works have modelled stock markets as time series and have studied the evolution of the assets over time, trying to predict future prices [9,6,12]. However, price evolution in financial markets is not depending on probabilities of specific outcomes.

From the point of view of the use of techniques based on CBR for predicting future prices whatever the context of the problem, few works have been carried out. CBR-based techniques have also been used in a broad range of applications [4,10] but have hardly been applied for the prediction of future prices in financial markets. However, sports betting markets has a critical temporal component which differentiates them from other financial markets and therefore, a CBR approach could identify similar price movement patterns that are repeated in different events, under similar circumstances.

The agent paradigm has been widely used in other competitive scenarios such as the Trading Agent Competition [14], a Fishmarket [13] or artificial stock markets [7]. In these scenarios agents can learn, collaborate and evolve their different strategies in order to compete with other participants. Sports betting markets are scenarios which also require competitiveness and in which the use of agent-based approaches can be very helpful in order to improve the performance of human participants.

3 Sports Betting Markets

Sports betting markets are speculative scenarios about sporting events, where participants exchange assets regarding a specific outcome of the event. For a specific event, there are several markets regarding the winner, the final score, the number of goals or points, handicap markets, etc. Each market has n possible outcomes and each bet can be seen as an $n-way$ bet. The prices of the exchanged assets are related to the probability of the specific outcome happening. Price and implicit probability of an outcome are related by $price = 1/probability$, where probability is represented from 0 to 1.

Trades are made between participants at a given price because they have different expectations about the event. In sports betting markets, users can bet on an outcome (win if this is the final outcome) or against it (win if any of the other outcomes is the final one). Betting on a specific outcome is called *back* and betting against it is called *lay*.

The bookmaker is in charge of receiving the offers of the traders. If a *back* offer and a *lay* offer are compatabile in terms of price and stake (full or partial), the bookmaker matches both offers. If a received offer is not compatible it remains waiting until a compatible offer is received, or until it is deleted by the user or because the market is closed. Therefore, the bookmaker also maintains a list of waiting offers, and continuously shows the best *back* and *lay* prices of these waiting offers in order to allow interested users make offers at these prices.

Let us suppose that user $Alice$ wants to bet μ units on placing a *back* bet on a selection (an individual, a team, horse, etc.). This user is betting that the selection will win. $Alice$ can accept the best waiting *lay* offer (the *lay* offer which price is the lowest one) or can choose his own price ρ. When user Bob wants to place a *lay* bet on this selection (against the selection of $Alice$, that is, the individual will not win), he can also choose his own price or accept the offer of $Alice$.

If *Bob* accepts *Alice*'s offer, *Bob* is placing a *lay* bet on the seletion at price ρ. When the event is over, if *Alice* wins the bet, *Bob* has to pay $\rho - 1$ units for each unit bet on. If *Bob* wins the bet, he keeps the μ units of *Alice*:

$$\text{profit(Alice)} = \begin{cases} \mu^*(\rho\text{-}1) & \text{if Alice wins the bet} \\ -\mu & \text{if Alice loses the bet} \end{cases}$$

$$\text{profit(Bob)} = \begin{cases} -\mu^*(\rho\text{-}1) & \text{if Alice wins the bet} \\ \mu & \text{if Alice loses the bet} \end{cases}$$

3.1 Trading on Sports Betting Markets

Sports betting markets are traditionally used for eliciting the probabilities of final outcomes before the event is started, but can also be approached while the event is being played. As an example, when a soccer match starts it has associated a price ρ for the outcome *0-0 as a final score* because users estimate that the probability of this outcome is ϕ, where $\rho = 1/\phi$. As the match draws to a close and the score remains 0-0, the probability increases and therefore, the price decreases. As sport betting markets have a short or very short duration, the prices change rapidly during the event.

From the point of view of a sports betting market trader, similar to an economic trader, the underlying sporting event and its final outcome is not important, because its goal is to make profits whatever the final outcome is.

In this approach, a sports betting market is seen as a market where the current price of a specific outcome is going to change over time and therefore, the importance relies on detecting if the price will move up or down, how much it will move up or down, when it is going to move and how fast. Therefore, the goal of a sports betting markets trader is to bet on one outcome at a higher price and to bet on the opposite outcome at a lower price.

A trader can make a bet by risking μ units whilst making a *back* bet at a price of ρ_1 for a specific outcome ω. As explained in Section 3, if the bet is finally won, the trader wins $\mu * (\rho_1 - 1)$ units, and loses the μ units if the other outcome is the final one. This trader can make profits if he covers all the bases by betting on the opposite outcome when prices change. In this example the trader can bet μ on the *lay* side at a price of ρ_2. When ending the event, if the final outcome is ω, the trader will win $\mu * (\rho_1 - 1)$ because his first bet has won, but he loses $\mu * (\rho_2 - 1)$ units bet on the second one.
As we can observe, if the price of the *back* bet (ρ_1) is higher than the price of the *lay* bet (ρ_2) the resulting profit will be positive:

$$\mu * (\rho_1 - 1) - \mu * (\rho_2 - 1) = \mu * (\rho_1 - \rho_2)$$

Nevertheless, if the ω outcome is not the final one, the trader will not lose any units, because he loses the μ units risked in the first bet but wins μ units from the second one. Thus, regardless whether ω is the final outcome or not, the trader will not lose any unit:

$$profit = \begin{cases} \mu^*(\rho_1\text{-}\rho_2) & \text{if } \omega \text{ is the final outcome} \\ 0 & \text{otherwise} \end{cases}$$

The same operations can also be made in the inverse order and also the trader can also split the profits between the different outcomes. Due to space restrictions, we do not detail these operations.

As we can observe, the difference in the prices of both trades indicates the amount of profit. Detecting price tendencies and therefore, predicting future prices is the key issue for making profits in these markets.

4 The Market Model

Sports betting markets represent a multilateral market model in which traders send their bets at their own price to the mediator who matches compatible bets. Orders compete for the best *back* and *lay* offers. Therefore, the offers which cannot be matched remain waiting until they can be matched or are cancelled. One of the tasks of the bookmaker is to also show at anytime the best *back* and *lay* prices of these waiting offers. For a specific market there is a list of all the currently *back* and *lay* bids currently waiting, ordered from the highest prices to the lowest.

In this work we use Betfair[1] as the sport betting market studied. Betfair is the world's biggest prediction exchange. According to [1] Betfair processes more than 6 million transactions in an average day (more than all of the European stock exchanges combined). Betfair is based on the New York Stock Exchange model and allows punters to bet at odds set by other punters rather than the bookmaker.

The Continous Double Auction (CDA) is a typical institution of real world exchange markets, such as financial assets, foreign exchange, energy, etc. In this institution, buyers and sellers place their offers at anytime. When a participant accepts a buy or a sell offer, a transaction is made. For modeling a sports betting market, we define a CDA institution where agents can interact for obtaining information of the market at a given moment and also, for placing bets.

In our model, Betfair acts as the mediator between users, matching the compatible bets and showing the best *back* and *lay* prices at a given time. As a wrapper of Betfair, we define the *bookmaker* agent which acts as a gateway between Betfair and the agents. Therefore, when a agent wants to request prices or wants to place offers, it needs to communicate with the *bookmaker* agent. If agents send offers, these will be matched by Betfair or will be queued in the waiting offers queue. If agents are requesting the current prices, the *bookmaker* agent will retrieve them by accessing Betfair. Therefore, from the point of view of other users they do not know if they are trading with humans or agents.

5 The CBR Agent

We can observe that in a tennis match each time the favorite player wins a set, the price of the outcome *winner of the match* gets lower, due to the fact that

[1] `http://www.betfair.com`

the probability of winning is higher. Similarly, the prices of the the *winner* of a basketball match decrease if the favourite team increases the score difference over time. In light of these repeated patterns and the similar movements of prices regarding the state of a sporting event, we propose an agent with an underlying CBR system which captures some features of a current event and finds similarities with other past events. Then, observing the price evolution in these historical events, the agent will be able to predict the more accurate future prices depending on what happens during the event from now on. From now on, we reference this agent as CBR-agent.

We can summarize the tasks that the CBR-agent carries out as follows:

Data acquisition and creation of the case base: the first step is the data acquisition according the requirements of the problem. The CBR-agent interacts with the *bookmaker* agent in order to receive information about sport events. Then, after a data filter process (for excluding samples which may not reveal a real probability at a given moment), the CBR-agent creates the case base which will be used in the CBR cycle.

CBR cycle: once the CBR-agent has created the case base, this is used for solving an unknown problem (in our case predict future prices) given similar past problems. The CBR-agent interacts with the *bookmaker* agent to obtain the information of a sporting event (unknown problem), then the CBR cycle measures the similarity between this problem and a similar past one (one or more). Then, according the future prices of the past problem, future prices are predicted for the unknown problem by adapting the solution of the past one.

In following sections we show an example about prediction in soccer events. By means of this example we detail the processes described above and then we show some results about the system accuracy prediction.

5.1 Experiments

We have carried out the experiments using a real data set from soccer matches played in the 2008-2009 season in the Barclays Premier League. We analyzed this competition because it is one of the most important soccer leagues in the world. Moreover, the large amount of traders that exchange bets at these events means that each event has very high liquidity, and this is important for obtaining more reliable results.

We focus on the price prediction for markets that are the under/over 2.5 goals. These markets show the probability assessed by the participants for scoring in a soccer event less than 2.5 goals (0, 1, or 2) or more than 2.5 goals (3 or more), respectively. We want to learn how the prices evolve depending on the current features of a current game.

5.2 Data Acquisition and Creation of the Case Base

There are several markets that are directly related to the under/over 2.5 goals, whose price evolution should be proportional to the markets studied since they refer to similar final outcomes. These markets are, for example, the markets regarding the exact final score or other markets regarding under/over goals (such as 1.5, 3.5, or 4.5 goals). Other markets are indirectly related to under/over 2.5 goals such as the match odds markets (the winner of the game or a draw) or the next-goal-minute markets. Although the price evolution in these markets may not be proportional to the studied markets (under/over 2.5 goals), it is related to the prices of the markets studied. Finally, other markets such as the number of yellow cards in the game, the number of corners, or the injury time of the event are not related to the studied markets.

Although all the soccer events are completelly different from each other (different players, different teams, weather conditions, dates, and so on), if we consider the prices of the outcomes of different markets, we can find some similarities between an unknown event and a past one. In the example presented in this paper, we define the state of a soccer event at a specific moment according to the next properties:

– The exact moment of the game (in minutes).
– The current score of the game at that particular moment.
– The prices of the under/over 2.5 markets. These show the *back* and *lay* prices for both the under and the over 2.5 goals outcomes.
– Match odds prices. These show the *back* and *lay* prices for the home wins, visitor wins and draw wins.

Taking into account these properties, we can find similarities for two different soccer events and predict future prices.

Every 60 seconds the CBR-agent requests the current values of these properties from the *bookmaker* agent, and then, the CBR-agent stores this information as a sample. Thus, we obtain some samples for a single soccer event. A sample is a quintuple which is described as $\langle m, s, h, v, d, u, o \rangle$ where:

– $0 \leq m \leq 45$ represents the minute of the game. For reasons of simplicity, we study the price evolution in the first half of the event (45 minutes). Thus, this component is an integer.
– $s \in \{0\text{-}0, 1\text{-}0, 0\text{-}1, 1\text{-}1\}$ represents the current score of the game. For reasons of simplicity, we only take into account soccer events that have these current scores.
– h, v, and d are respectivelly the prices refering to the home wins, the visitor wins, and the draw wins from the match odds market.
– u and o are the prices refering to the under outcome and the over outcome from the under/over 2.5 goals market.

Each h, v, d, u and o has two real values $\langle b, l \rangle$ that represent the *back* and *lay* prices for the specific outcome. Each *back* price must be lower than the corresponding *lay* price at any given moment. Otherwise, there would be a possible trade.

The maximum difference allowed between *back* and *lay* prices is represented by ϵ. Therefore, each pair of *back* and *lay* values $\langle b, l \rangle$ fulfills $b < l \leq b + \epsilon$. This ϵ is a specific threshold that we use in order not to consider samples in which at least one pair of *back* and *lay* prices differs more than ϵ.

Since samples are taken every 60 seconds, we create a sequence of samples $x_1, x_2, x_3, \ldots x_n$ for each different event. To simplify notation, if $x_i = \langle m, s, h, v, d, u, o \rangle$, we write m_i to refer to m, and similarly for other components of x_i. It is assumed that the samples are ordered by time, i.e., $m_i < m_j$ whenever $i < j$.

In order to predict the prices of an unknown event in the next δ minutes, we need to find similarities with other past events whose prices are known both at a specific time and after δ minutes. Therefore, we, we need to represent information as a problem description and its solution. In our case, the problem description is each one of the stored samples of a single event and the solution is the state of this event after δ minutes. In the example presented in this paper, we make predictions for the next, 5, 10 and 15 minutes ($\delta = \{1, 5, 10, 15\}$).

Thus, given two different samples of the same event $x_i = \langle m_i, s_i, h_i, v_i, d_i, u_i, o_i \rangle$ and $x_j = \langle m_j, s_j, h_j, v_j, d_j, u_j, o_j \rangle$, such that $m_j - m_i = \delta$, we define a case of the case base as:

$$c^\delta = \langle m_i, s_i, h_i, v_i, d_i, u_i, o_i, d(u_i, u_j), d(o_i, o_j) \rangle$$

Each case represents the information of the event in the moment m_i (initial moment) and the information regarding the event after δ minutes, which in our example is the information regarding the under/over outcomes: $d(u_i, u_j)$ and $d(o_i, o_j)$. We define an operation d on pairs of *back* and *lay* values as follows: $d(\langle b_1, l_1 \rangle, \langle b_2, l_2 \rangle) = \langle \mid b_1 - b_2 \mid, |y_1 - y_2| \rangle$. Thus, $d(u_i, u_j)$ and $d(o_i, o_j)$ represent the *back* and *lay* price variations for the under/over markets.

All cases c^δ of all the events that we have two samples x_i and x_j such that $m_j - m_i = \delta$, allow us to create a case base $C^\delta = \{c^\delta \mid c^\delta \text{ is defined}\}$. This case base stores cases of different events, but the information represented in a single case obvioulsly refers to a specific event.

5.3 The CBR Cycle

The case base for a specific δ represents the information of events and their *back* and *lay* prices for the under/over markets in the next δ minutes. Therefore, given C^δ and an input problem $x = \langle m, s, h, v, d, u, o \rangle$, the CBR cycle predicts the *back* and *lay* prices for the under/over markets in the next δ minutes.

The inference process of the CBR system can be summarized in the following steps:

Step 1. Retrieve the cases whose score components are the same and also whose *back* and *lay* prices for the under/over markets in the initial moment are the most similar to the prices of the input problem. For example, given an input problem $x = \langle m, s, h, v, d, \langle b_u, l_u \rangle, \langle b_o, l_o \rangle \rangle$ and a case of the case base $c_r = \langle m_r, s_r, h_r, v_r, d_r, \langle b'_u, l'_u \rangle, \langle b'_o, l'_o \rangle, u^\delta, o^\delta \rangle$, c_r is retrieved if the score is the same than the input problem and if its values are not different from the values of the input problem by more than a threshold ω:

$$s = s_r \wedge \mid b_u\text{-}b'_u \mid \leq \omega \wedge \mid l_u\text{-}l'_u \mid \leq \omega \wedge \mid b_o\text{-}b'_o \mid \leq \omega \wedge \mid l_o\text{-}l'_o \mid \leq \omega$$

Step 2. From all cases retrieved in Step 1, we select those whose time component is the most similar. For example, given an input problem $x = \langle m, s, h, v, d, u, o \rangle$

and a case of the case base $c_r = \langle m_r, s_r, h_r, v_r, d_r, u_r, o_r, u^\delta, o^\delta \rangle$, we select those cases where $\mid m\text{-}m_r \mid < \pi$.

Step 3. From each case c_r selected in Step 2, where:

$$c_r = \langle m_r, s_r, h_r, v_r, d_r, \langle b_u, l_u \rangle, \langle b_o, l_o \rangle, \langle b'_u, l'_u \rangle, \langle b'_o, l'_o \rangle \rangle$$

for $r = \{1, 2 \ldots R\}$, being R the number of selected cases, we calculate price evolution for the specific price (*back*,*lay*) and market (under,over) as follows:

$$e(b,u) = (b_u - b'_u);\ e(l,u) = (l_u - l'_u);\ e(b,o) = (b_o - b'_o);\ e(l,o) = (l_o - l'_o)$$

If a price evolution is positive it means that the specific price is going to decrease in δ minutes; if it is negative, the price will increase δ minutes.

Then, we calculate an average price evolution from all cases $c_1, c_2 \ldots c_R$, for the specific price (*back*,*lay*) and market (under,over):

$$A(b,u) = \frac{1}{R}\sum_{r=1}^{R} e(b,u)_r;\ \ A(l,u) = \frac{1}{R}\sum_{r=1}^{R} e(l,u)_r$$

$$A(b,o) = \frac{1}{R}\sum_{r=1}^{R} e(b,o)_r;\ \ A(l,o) = \frac{1}{R}\sum_{r=1}^{R} e(l,o)_r$$

Consider an input problem $x = \langle m, s, h, v, d, \langle b_u, l_u \rangle \langle b_o, l_o \rangle \rangle$ and price evolutions for the specific price (*back*,*lay*) and market (under,over): $A(b,u), A(l,u)$, $A(b,o), A(l,o)$. Then, the predicted *back* and *lay* prices for the under/over markets would be:

$$predicted(p,k) = p_k + A(p,k)$$

for each $p = \{b, l\}$ and $k = \{u, o\}$.

Step 4. If the predicted prices are similar to the real ones, the case is then retained in the case base. If one of the four predicted prices is different by more than a specified threshold, the case is not retained, assuming that this case may be an anomalous case. If so, storing it could decrease the prediction accuracy of the entire system. We dynamically change this threshold as the size of the case base increases.

In Step 1, we used a ω threshold of 0.05 for $\delta = 1$ (predictions for the next minute). In other experiments ($\delta = \{2, 5, 10\}$) we used a threshold of 0.20. In Step 2 we used a π value of 1. That is, we took into account only samples whose time component was not greater or lower than the time component of the input problem by more than 1 minute. In Step 1, if no case is retrieved, we increase the threshold to a maximum of twice the initial threshold. If increasing the threshold is not enough to retrive a case, we then consider the values of home, visitor, and draw in order to find similarities with these components. In Step 2, if no case is selected according to this restriction, we increase the π value until we find a case.

6 Results

In Table 1 we can see the price prediction accuracy depending on the number of past events observed. The agent predicts the future prices for 100 sporting events. The table shows in how many of these 100 cases the predicted price was the real one, with some error rate.

We can see that as the agent increases the number of past observed cases, the accuracy of the future price prediction is also increased. We can conclude that future events follow price movement patterns similar to past events. We can see that with 250 cases in the case base, the precision accuracy is around the 90% for an error rate of 0.05. Moreover, for predictions in the next minute this accuracy is almost 100% for the same error rate. Prices are represented in cents and are usually placed between 1.01 and 3 in these markets, thus, this accuracy is high, and probably gets higher if we increase the number of observed cases.

Table 1. Number of cases in wich the predicted price is the real one with error rates

Error Rate	50 cases				150 cases				250 cases			
	Under		Over		Under		Over		Under		Over	
	B	L	B	L	B	L	B	L	B	L	B	L
Prediction for the next minute												
±0.02	73	63	71	64	81	75	72	71	83	78	69	73
±0.03	85	78	81	77	89	85	85	80	91	88	83	82
±0.05	93	93	95	90	96	95	95	91	97	97	95	94
±0.1	100	99	99	99	99	99	99	99	100	100	99	99
Prediction for the next 5 minutes												
±0.02	44	46	42	44	62	60	50	46	68	68	46	49
±0.03	52	66	56	56	72	70	54	62	74	77	65	71
±0.05	74	80	72	76	82	86	78	80	94	87	94	84
±0.1	96	94	88	96	100	96	88	92	100	100	100	97
Prediction for the next 10 minutes												
±0.02	32	41	22	32	54	38	32	35	63	62	44	44
±0.03	43	49	32	41	68	49	43	51	69	67	62	64
±0.05	59	65	41	49	84	73	65	73	91	88	84	80
±0.1	84	86	70	84	100	89	79	95	100	98	96	97
Prediction for the next 15 minutes												
±0.02	43	47	17	23	57	57	43	37	64	66	45	42
±0.03	53	50	37	30	73	67	50	43	74	70	61	64
±0.05	67	67	40	47	91	80	60	57	94	89	87	86
±0.1	97	90	63	77	100	93	73	80	100	98	95	97

Predictions for the next 5, 10 or 15 minutes have a similar accuracy, but predictions for the next minute are slightly more accurate. We can observe that the accuracy is around 70% with an error rate of 0.03. If the error rate is 0.05 the accuracy is around 95%.

Apart from the accuracy precision, it is important to check the price direction accuracy when the predictions are made very quick. In order to draw a price evolution over time we should be able to successfully predict if price increases or decreases at the short-term. In Table 2 we can see a combination of prediction intervals and success rates for four different soccer events during 45 minutes. In this table we show in how many minutes the price direction in the next minute is successfully predicted for the *back* price of the under market.

Table 2. Success rates for price direction prediction in the next minute

Event	Price Direction Accuracy [±0.03]	%
Match 1	34	0.77
Match 2	30	0.68
Match 3	39	0.89
Match 4	30	0.68

We can observe that the market behaviour can be learned due to the success rates for each match are quite accurate. For each different event, the CBR-agent successfully predicts the price direction in practically the 70% of the minutes or even more. Therefore a price evolution from minute 0 to 45 should be quite similar to the real one.

7 Conclusions

In this paper we have seen how sports betting markets can be approached as trading scenarios for making profits regardless of the final outcome of the event. The approach seen in this paper is different from the approaches in which prediction markets have been studied. In these markets, it is important to predict future price in order to identify the best trading decisions.

We focus on predicting future prices by means of detecting future price movements. As the price is related to the probability, and this probability changes as the sporting event is being played, price movements can follow patterns in different events with similar circumstances. Due to the short duration of sporting events, sports betting markets display quick exchanges. Our aim is to identify if the price will move up or down, how much it will move up or down, when it is going to move and how fast can be learned from past events.

From these features, we present a prediction system based on a CBR approach. Observing events, we develop a CBR-agent which is able to find the similarities of a unknown event with other historical ones by using CBR. Therefore, the agent predicts future prices according to its reasoning.

The accuracy of the predictions has demonstated that despite each event being different, under similar circumstances some price movement patterns are repeated. The agent is able to learn these patterns and identifies them in other uncertain events for predicting future prices.

Although we have presented our experiments using a specific sport, other sports should also repeat price movement patterns under the same sporting event circumstances. Thus, the results could be applied to other sports. We also plan to apply CBR prediction for other sport betting markets.

Another area for future work is to identify the most influential factors on the price movements. We can retrieve cases according to different similarity functions and prove which technique make more accurate predictions. The multilateral market model designed should allow us to compare heterogeneous agents with different trading strategies.

References

1. Betfair Corporate, http://www.betfaircorporate.com
2. Chen, Y., Goel, S., Pennock, D.: Pricing combinatorial markets for tournaments. In: STOC '08: Proceedings of the 40th Annual ACM Symposium on Theory of Computing, pp. 305–314. ACM, New York (2008)
3. Debnath, S., Pennock, D.M., Giles, C.L., Lawrence, S.: Information incorporation in online in-game sports betting markets (2003)
4. Gayer, I.G.G., Lieberman, O.: Rule-based and case-based reasoning in housing prices (2004)
5. Guo, M., Pennock, D.: Combinatorial prediction markets for event hierarchies. In: Proceedings of The 8th International Conference on Autonomous Agents and Multiagent Systems, pp. 201–208 (2009)
6. Huang, W., Lai, K., Nakamori, Y., Wang, S.: Forecasting foreign exchange rates with artificial neural networks: A review. International Journal of Information Technology and Decision Making 3(1), 145–165 (2004)
7. LeBaron, B.: Agent based computational finance: Suggested readings and early research. Journal of Economic Dynamics and Control (1998)
8. Mantaras, R.L.D., McSherry, D., Bridge, D., Leake, D., Smyth, B., Craw, S., Faltings, B., Maher, M., Lou, C., Forbus, M.C.K., Keane, M., Aamodt, A., Watson, I.: Retrieval, reuse, revision and retention in case-based reasoning. Knowl. Eng. Rev. 20(3), 215–240 (2005)
9. Moody, J.: Economic forecasting: Challenges and neural network solutions. In: Proceedings of the International Symposium on Artificial Neural Networks (1995)
10. Oh, K., Kim, T.: Financial market monitoring by case-based reasoning. Expert Syst. Appl. 32(3), 789–800 (2007)
11. Plott, C.: Markets as information gathering tools, pp. 1–15 (2000)
12. Raudys, S., Zliobaite, I.: The multi-agent system for prediction of financial time series. In: Rutkowski, L., Tadeusiewicz, R., Zadeh, L.A., Żurada, J.M. (eds.) ICAISC 2006. LNCS (LNAI), vol. 4029, pp. 653–662. Springer, Heidelberg (2006)
13. Rodriguez-Aguilar, J.A., Martin, F.J., Martn, F.J., Noriega, P., Sierra, C., Garcia, P.: Competitive scenarios for heterogeneous trading agents. In: Proceedings of the Second International Conference on Autonomous Agents (1998)
14. Wellman, M., Greenwald, A., Stone, P., Wurman, P.: The 2001 trading agent competition. IEEE Internet Computing 13, 935–941 (2000)