Predicting sports events from past results

Towards effective betting on football matches

Douwe Buursma
University of Twente
P.O. Box 217, 7500AE Enschede
The Netherlands
d.l.buursma@student.utwente.nl

ABSTRACT

A system for predicting the results of football matches that beats the bookmakers' odds is presented. The predictions for the matches are based on previous results of the teams involved.

Keywords

Prediction, odds, bookmakers, football, probability, profit, Weka, Bayesian networks, regression analysis.

1. INTRODUCTION

Betting on sports events has gained a lot of popularity recently and bookmakers and betting offices are now omnipresent at the internet. One of the most popular sports to bet on is football, because hundreds of football matches are played every week in professional competitions around the globe.

The method is always the same, the bettor wagers an amount of money on a bet of their choice. If their prediction turns out to be correct they win an amount of money, equal to the odds of the bet, per unit of money invested. For instance, if a bettor wagers 3 units of money on a bet with odds 1.75, they will win 3*1.75=5.25, yielding a net profit of 2.25. If the prediction is incorrect the bettor's wager is lost. The odds bookmakers use for their bets are based on probability. If the probability of a certain outcome showing up is 0.2, the corresponding odds would be 1/0.2 = 5, meaning every unit bet will result in 2 units returned if the bet is won. Obviously, bookmakers like to make a little profit, so when they believe the probability of a certain outcome showing up is 0.2, they will not use the statistically fair odds of 5, but will rather adjust this to for instance 4.5.

Because of this "unfairness" by the bookmakers, making profit by betting on football results is not all that easy. One would so to speak have to "beat" the bookmakers' odds. In order to do this, accurate probabilities for a home win, a draw or an away win must be calculated for each match. Section 2 and 3 give a detailed description of how this is done.

1.1 Related Research

Lots of research on this topic has already been done. In 2003, Buchdahl [2] investigated the quality of bookmakers' odds in all kinds of sports, including football. He concluded that, when enough data is used for training a prediction system, even a small feature set will suffice to beat the bookmakers' odds in football. With a larger and more intelligent feature set, it should be possible to predict football matches even more accurately and therefore gain a higher profit by betting. Goddard [5], in 2004, did use a much bigger feature set for his football match predictions. This feature set includes the geographical distance between the teams involved and their average spectator attendances. However, the profit of their system was a lot lower than the one by Buchdahl. We believe the main reason for this is that they simply bet an amount of money on the most probable outcome for each match. Reasonable though this may seem, it is not too hard to understand that, because of bookmakers' unfairness, it may sometimes be better not to bet on a match at all. In our investigation an attempt will be made to accurately choose which bets to take on in order to maximize the profit.

1.2 Research Questions

The main question of this research is:

- How can we effectively bet on football matches by using features from past results?

To answer this main question, several sub-questions will have to be answered.

- What features are important when predicting the outcome of a football match?
- How can we calculate the probabilities for the outcomes of a match from these features?
- Knowing these probabilities, what bets should we take on to maximize the profit?

1.3 Approach

1.3.1 Database and setup

To predict football matches accurately several things are needed. First of all, a set of features containing the most valuable information about previous matches must be collected. To do this, a database containing data from the past 15 years for the Dutch football competition is used. This database is available on http://www.football-data.co.uk/. The database contains all full-time and half-time score lines and the bookmaker odds for most of the matches. From these data lots of features, such as *number of goals scored in the last 7 matches*; *number of home wins in the last 10 matches* etc., can be extracted. In Section 2 it is described in more detail what features are used and how they are collected.

After collecting a vast number of features, we will then let Weka (www.cs.waikato.ac.nz/ml/weka/) [11] determine what features are important. Weka is a tool that uses machine learning algorithms in order to calculate the correlation between certain data. Some of the most important machine learning algorithms used in this investigation are the Bayesian network[1] and forms of linear and logistic regression[10]. Basically what these algorithms do, is calculate how a given vector of outcomes is determined by a vector of numbers corresponding to these outcomes. For our investigation we will

14th Twente Student Conference on IT, January 21st, 2011, Enschede, The Netherlands.

use these algorithms to calculate the correlation between the feature set that is obtained for each match and the outcome of the match. The outcome will be of the form “H”, “D” or “A”, meaning “home win”, “draw” and “away win” respectively. This analysis will, for all matches in the database, lead to a vector of percentages, indicating the probability of either a home win, a draw or an away win to show up. It then chooses the most likely outcome as its prediction and in the end it checks how many of the predicted outcomes were correct. By letting Weka run several different machine learning algorithms on the data we can select the one that performs, i.e. the one that yields the highest percentage of correctly predicted outcomes.

1.3.2 Training and testing

The process during which we let the machine learning algorithm calculate the probabilities for the matches is called *training*. But, since we also want to put our system to the test later on, we have to reserve a set of data that we will later use for this purpose. Naturally, we cannot use this set of data during the training phase as well, or else our system would become biased. The bigger the percentage of the data that is used for training, the better the system will fare, simply because the analysis can use more data. On the other hand, the bigger the percentage of the data that is used for testing, the more statistically reliable our test will be. Now, what would be a good way of splitting up our fifteen seasons of football matches? Weka offers a very good solution for this problem, namely a ten-fold cross validation. It splits the data into ten equal-sized portions and uses nine out of ten portions as training data and the last one as testing data. It repeats the process ten times, each time choosing a different portion as the testing data.

Then, the test phase will commence. During this phase a prediction for all matches will be made and this will be compared to the actual result. Our investigation focuses on betting, so instead of simply scoring the number of correct predictions, we value each correct prediction. This value corresponds to the bookmaker odds for this specific result of the match. The performance of a betting strategy is measured in the following way. Each correctly predicted match means we gain the number of units corresponding to the odds for that result; each match a bet is wagered on means we lose one unit. In other words, we lose one unit for an incorrect prediction and we gain the odds minus one unit for a correct prediction.

A naïve way to go about this would be to simply bet on the most probable outcome for every match. As described in Section 1.1, one can easily see that it is sometimes better not to bet at all on a certain match. Section 4 describes in full detail all the betting strategies used in this research and also lists the results.

2. FEATURE SET

2.1 Match history

An accurate feature set makes it a lot easier to predict the outcomes of matches. So, selecting relevant features is an important process. Features are characteristics of recent matches of the teams involved, but how far in history do we need to go in order to get the best predictions? To answer this question we set up a very basic set of features and then each time we changed the amount of history looked at and compared the results.

This initial set included the following features:

- Goals scored by home team in its latest x matches
- Goals scored by away team in its latest x matches
- Goals conceded by home team in its latest x matches
- Goals conceded by away team in its latest x matches
- Average number of points gained by home team in its latest x matches
- Average number of points gained by away team in its latest x matches

The x stands for the (variable) number of matches looked at. The first four features are pretty straightforward, the last two describe the points the home and away team gained in their latest matches. These are calculated as in the football competition itself, namely, 3 points for a win, 1 for a draw and 0 for a loss. The average over the latest x matches is taken. By importing the features in Weka and letting several machine learning algorithms classify the data as described in Section 1.3, a percentage of correctly predicted instances is given.

For several values of x (number of matches looked at) we compared the result. For two of the best classifiers (machine learning algorithms) the results are given in Table 1.

Table 1. Classification performance when considering x matches. The left column indicates the classifier used and the other columns show the percentages of correctly predicted matches

Class. /x -->	**7**	**10**	**15**	**20**	**30**
ClassViaReg	52,55%	53,53%	54,23%	54,86%	54,97%
MultiCC	52,59%	53,46%	54,39%	54,84%	54,92%

It can be seen that the percentage keeps going up when x becomes larger, so apparently the best feature set would be one in which, per team, all historic matches are used. Of course, since our database is limited and because the percentages do not go up very fast after x becomes larger than 20 we decided to stick to this number for the rest of our investigation. A full list of results can be found in Appendix A.1. Although our database consists of multiple seasons, they are put together seamlessly. So, when for a fixture at the beginning of the third season we look at the 20 latest matches for both teams, this will also include matches from the second season. Obviously, whenever one of the teams involved has not yet played 20 matches in the database, the match as a whole will be discarded and will not be used in either training or test data.

2.2 Classifiers

Now that an optimal number of matches to be considered has been found, we can move on to selecting the best possible classifier (machine learning algorithm). These will by means of a certain machine learning algorithm classify all matches as home wins, draws or away wins, depending on the features belonging to that match. During the previous test round a selection has already been made. Below is a list of seven classifiers including a short description of each:

- ClassificationViaRegression – This algorithm uses linear regression in order to predict the right class.
- MultiClassClassifier – This algorithm is a lot like ClassificationViaRegression, except that it uses logistic regression instead of linear regression.
- RotationForest – This algorithm uses a decision tree to predict the right class.
- LogitBoost – This is a boosting algorithm that also uses logistic regression.
- BayesNet – This algorithm uses Bayesian networks to predict the right class.

- NaiveBayes – This algorithm resembles BayesNet, except
- Home wins – This algorithm will, regardless of the feature set, always predict a home win.

In the previous section we have already seen that the first two perform best, using the given simple feature set. We now expand our feature set by a few more features and make several selections of them to see which classifier is best. Please note that the "home wins"-classifier is used merely as a reference. It can immediately be seen that this classifier performs worse than all the others.
As stated above, an expansion will be made to the existing feature set. Below is a new, numbered, list of features, from which the selections will be made.

- (1) Goals scored by home team in its latest 20 matches
- (2) Goals scored by away team in its latest 20 matches
- (3) Goals conceded by home team in its latest 20 matches
- (4) Goals conceded by away team in its latest 20 matches
- (5) Average number of points gained by home team in its latest 20 matches
- (6) Average number of points gained by away team in its latest 20 matches
- (7) Number of home matches won by home team in its latest 20 matches
- (8) Number of away matches won by away team in its latest 20 matches
- (9) Number of goals scored in home matches by home team in its latest 20 matches
- (10) Number of goals scored in away matches by away team in its latest 20 matches

Table 2 shows the performance of all the classifiers on certain feature selections. The best performing classifier is highlighted in each column.
It becomes quite clear from the table that the ClassificationViaRegression and the MultiClassClassifier-algorithms almost invariably also perform best on these feature selections. Since the former performs slightly better, we will use that algorithm for the rest of this investigation. One can already make few first predictions about which features are important based on this table, but we will leave that for section 2.3. A full overview of the results of all feature selections can be found in Appendix A.2.

2.3 Features

Now that a classifier has been selected and the number of matches to consider is also known, we have come to the most important part of this chapter, the actual selection of the features.
For the first part of this selection procedure we will use the exact same feature set we used in Section 2.2. Then, one at a time, we will leave features out. If the percentage drops after we left the feature out, the feature is deemed important and should be left in, if the percentage rises, the feature is discarded. Should the difference be marginal, e.g., smaller than 0,04%, we will decide to leave the feature in. It turns out that only features no. 7 and 8 make the percentage drop, so we will discard those features. The results for this can be found in Appendix A.3. This means we are left with a set of eight features.

The next part of the selection procedure concerns the one by one adding of new features. This time obviously, if the percentage rises after adding a feature, the feature will be left in. Should the percentage drop, the feature is discarded.

Table 2. Performance of the classifiers when using different selections of features

Classifier/features-->	1-10	1-6,9,10	1-6	1,2
ClassViaReg	55,00%	55,05%	54,86%	53,59%
MultiCC	54,86%	54,98%	54,84%	53,73%
RotationFor	53,61%	54,57%	54,04%	53,06%
LogitBoost	53,88%	54,62%	54,00%	53,30%
BayesNet	53,64%	54,55%	54,38%	53,59%
NaiveBayes	53,64%	54,38%	54,43%	53,37%
HomeWins	48,66%	48,66%	48,66%	48,66%

A list of all features to be added is given below:

- (11) Goal difference of home team in its latest 20 matches
- (12) Goal difference of away team in its latest 20 matches
- (13) Number of matches played at home by home team in its latest 20 matches
- (14) Number of matches played at home by away team in its latest 20 matches
- (15) Number of goals conceded in home matches by home team in its latest 20 matches
- (16) Number of goals conceded in away matches by away team in its latest 20 matches
- (17) Average strength of the home team's opponents in its latest 20 matches
- (18) Average strength of the away team's opponents in its latest 20 matches

These last two features, indicating the average strength of a team's opponents are calculated as follows: for all opponents in its last 20 matches the average number of points gained (this is equal to feature no. 5/6) is counted and the average of these 20 values is taken. Should an opponent not have played 20 matches before the given encounter yet, then the average is simply taken over the number of matches they have played.

Adding features 11 and 12 does not affect the percentage at all. This should not be too surprising, as they can simply be calculated from features 1,3 and 2,4 respectively. Adding features 13 up to 17 makes the percentage drop, so these features are discarded, but adding feature 18 makes the percentage go up to 55,08%, so this feature will be left in.
A full list of results can again be found in Appendix A.3.
The feature set is now complete, so we can move on to calculating the probabilities for the outcomes of the matches.

3. CALCULATING PROBABILITIES

For predicting the outcomes of the matches classifiers must use a certain formula. This formula can also be used in order to calculate probabilities for these matches. Our classifier, the ClassifierViaRegression creates three different formulas, one for each possible result (home win, draw or away win). For the classification process the outcomes of these formulas are then compared and the one with the highest value is selected. For our investigation we are interested in probabilities for a certain match result, so the outcomes of these formulas have to be normalized, i.e. made to sum up to 1. In order to do so we simply take each individual outcome and divide it by the sum of the three outcomes. This is described in further detail, and along with the formulas created by the Weka classifier, in Appendix B.1.

These probabilities have to be compared to the odds of the bookmakers. As described in Section 1, bookmaker odds can easily be transformed into probabilities simply by taking their reciprocal value. In Section 4 several betting strategies based on these probabilities are tested.

4. TESTING AND RESULTS

Now that the probabilities for the outcomes of all the matches are known, the betting strategies can be evaluated. As stated in Section 1.3.2, the naïve way to go about this would be to simply bet on the outcome with the highest probability. But, it is also possible to compare the obtained probabilities to the ones given by the bookmakers' odds and to only bet if your probability is higher. Several different strategies, including a basic strategy which simply bets on a home win every match, are compared. The betting strategies used can be divided into four classes. The first(I) and second(II) class both contain only one strategy, namely, the above described "naïve" strategy and the one that bets on home wins respectively. The third(III) class contains strategies that will only bet if the obtained probability is more than a certain percentage, which we will call *y*, higher than the probability of the bookmakers. These strategies will be indicated by III-*y*. The fourth(IV) and final class contains strategies that will only bet if the obtained probability is more than a certain factor, which we will call *z*, higher than the probability of the bookmakers. These strategies will be indicated by IV-*z*. The basic case for both the third and fourth strategy is the one that bets on an outcome simply if the obtained probability is higher than the bookmakers' probability. This strategy can be called either III-*0* or IV-*1*.

The bookmakers' odds used for this investigation are those supplied by Bet365 (www.bet365.com). Unfortunately, in our database, these odds are only available for the last 7 seasons, so from 2003-2004 onwards. We will therefore evaluate the strategies on these data only.

Table 3 shows the most important results from these evaluations. A full results table is available in Appendix B.2.

It can be seen from the table that strategies from classes III and IV fare best, with strategy IV-1.16 being the overall winner. It can easily be explained that the best class IV-strategy beats the best class III-strategy. Since we are talking about probabilities a deviation of 0.08 on a probability of 0.2 is an entirely different thing from a deviation of 0.08 on a probability of 0.8. By using a factor instead of a fixed deviation, the deviation is automatically scaled down or up.

In the short run, strategy I can keep up with strategies from those classes, but in the long run, the latter systems clearly win. Strategy II is, as expected, simply hopeless.

The numbers in the table indicate the overall profit when a certain betting strategy is used over a course of time. A placed bet always costs 1 unit, whereas the winnings, when a bet was correct, are equal to the corresponding bookmakers' odds.

Table 3. Performance of strategies from different classes

Strategy	last ½ season	last season	all 7 seasons
I	4.41	5.44	-25.784
II	-11.13	-17.95	-119.716
III-0 or IV-1	-0.99	-4.86	12.486
III-0.06	4.09	6.42	45.177
III-0.08	1.15	-1.32	48.214
III-0.10	1.9	-0.32	40.144
IV-1.12	-2.41	-0.7	31.414
IV-1.16	3.68	4.51	56.504
IV-1.20	0.9	-1.57	40.82

5. CONCLUSIONS

Although the predicting system did not do quite as well as hoped beforehand, some interesting conclusions can be drawn from this investigation. First of all, the selection of the feature set led to a few important insights. Apparently the performance of the classifier kept improving the bigger we made the set of recent matches that was looked at. Whereas one would expect the optimal number of matches to lie around 10, it has been shown that when this number is 20, 30 or even 75 the performance keeps improving. Another interesting fact about the feature selection is that the history of opponents of the home team does not seem to play as important a role as the history of opponents of the away team.

The performance of the chosen classifier was at never much higher than a mere 55%. Nevertheless it was shown in Section 4 that, with the right betting strategy, this can in the long run lead to a profit at the bookmakers. This strategy turned out to have to be one that would only bet on a match when the probability given by the classifier was higher than the probability given by the bookmakers' odds.

The major conclusion, though, remains that football matches will always be very hard to predict.

6. RECOMMENDATIONS AND FUTURE WORK

The system presented in this research can, albeit not by much, beat the odds of the bookmakers. There are however a lot of things that can be improved on in the future. For this investigation all that was taken into account was the number of goals scored in matches. From this we could of course extract the winner of each match and the number of points gained by teams. But there are several other things important when predicting football matches: all bookings (yellow and red cards) during a match, the players each team has, their managers, the amount of money the team can spend on new players each year up to perhaps even the capacity of their home grounds. Using all or even some of these data will, however, also make the investigation a lot more complex.

Predicting systems created for football matches could in the same way be created for different branches of sports. Basketball, baseball and ice hockey for instance, are very similar to football. It is simply a matter of selecting the right features.

7. ACKNOWLEDGEMENTS
First of all I would like to thank Maarten Fokkinga and all track members for their help during the whole project. I would also like to thank Mannes Poel for his help and assistance with the Weka Toolkit.

8. REFERENCES

[1] Ben-Gal I., 2007. Bayesian Networks, in Ruggeri F., Faltin F. & Kenett R., Encyclopedia of Statistics in Quality & Reliability, Wiley & Sons. http://www.eng.tau.ac.il/~bengal/BN.pdf

[2] Buchdal, J., 2003. Rating Systems For Fixed Odds Football Match Prediction, from Fixed Odds Sports Betting: The Essential Guide. http://www.football-data.co.uk/ratings.pdf.

[3] Carmichael, F. and Thomas, D., 2005. Home-field effect and team performance: evidence from English premiership football. Journal of Sports Economics. v6 i3. 264-281.

[4] Goddard, J., 2003. Modelling Football Match Results and the Efficiency of Fixed-Odds Betting, Working Paper, Department of Economics, Swansea University.

[5] Goddard, J. and Asimakopoulos, I., 2004. Forecasting football results and the efficiency of fixed-odds betting. Journal of Forecasting. v23. 51-66.

[6] Halicioglu, F., 2005. Forecasting the professional team sporting events: evidence from Euro 2000 and 2004 football tournaments, in: Fifth International Conference on Sports and Culture: Economic, Management and Marketing Aspects.

[7] Joseph, A., Fenton, N.E., Neil, M., 2005. Predicting football results using Bayesian nets and other machine learning techniques (version with full set of models and data). http://www.dcs.qmw.ac.uk/~norman/papers/Spurs-2.pdf.

[8] Koning, R.H., 2000. Balance in competition in Dutch soccer. The Statistician. v49 i3. 419-431.

[9] Rue, H., Salvesen, Ø., 1997. Prediction and retrospective analysis of soccer matches in a league, preprint Statistics 10.

[10] Sykes, A.O., 1992. An introduction to Regression Analysis. http://www.law.uchicago.edu/files/files/20.Sykes_.Regression.pdf

[11] Witten I.H. and Frank E., 2000. Data Mining: Practical Machine Learning, Tools and Techniques with Java Implementations, Chapter 8. http://www.pami.sjtu.edu.cn/people/sqwang/Tutorial.pdf

APPENDIX

A. FEATURE SELECTION

In this appendix we present the full results of all evaluations done in Section 2.

A.1 Match History

Section 2.1 dealt with determining the best number of matches to consider for the feature set. Below, in table 4 is a comprehensive list of results, showing basically that the larger the number of matches considered, the better the classifiers perform. But it can also be seen that the performance does not improve by much when increasing the number of matches further than 20. Therefore this number is chosen as the "optimal" value.

Table 4. Classification performance when considering *x* matches

Class. /*x* -->	**4**	**5**	**7**	**9**	**10**
ClassViaReg	51,81%	51,88%	52,55%	53,64%	53,53%
MultiCC	51,81%	51,88%	52,59%	53,42%	53,46%
RotationFor	50,47%	51,11%	51,69%	51,82%	53,00%
LogitBoost	51,00%	51,63%	52,37%	52,62%	52,55%
BayesNet	49,69%	50,80%	51,80%	52,55%	52,66%
NaiveBayes	51,01%	51,47%	52,14%	53,05%	53,07%
Class. /*x* -->	**12**	**15**	**20**	**30**	**75**
ClassViaReg	54,17%	54,23%	54,86%	54,97%	55,14%
MultiCC	54,24%	54,39%	54,84%	54,92%	55,46%
RotationFor	53,50%	53,27%	54,04%	54,04%	54,68%
LogitBoost	52,96%	53,53%	54,00%	54,39%	55,30%
BayesNet	53,61%	53,58%	54,38%	54,11%	55,24%
NaiveBayes	53,36%	53,79%	54,43%	54,82%	55,02%

A.2 Classifier selection

Section 2.2 subsequently deals with choosing the best classifier. Table 5 shows the performance of all classifiers on all chosen feature selections. The feature list is also given below.

- (1) Goals scored by home team in its latest 20 matches
- (2) Goals scored by away team in its latest 20 matches
- (3) Goals conceded by home team in its latest 20 matches
- (4) Goals conceded by away team in its latest 20 matches
- (5) Average number of points gained by home team in its latest 20 matches
- (6) Average number of points gained by away team in its latest 20 matches
- (7) Number of home matches won by home team in its latest 20 matches
- (8) Number of away matches won by away team in its latest 20 matches
- (9) Number of goals scored in home matches by home team in its latest 20 matches
- (10) Number of goals scored in away matches by away team in its latest 20 matches

Table 5. Performance of all classifiers when using different selections of features

Classifier/features-->	**1-10**	**1-6,9,10**	**1-4**	**1,2**
ClassViaReg	55,00%	55,05%	54,96%	53,59%
MultiCC	54,86%	54,98%	54,88%	53,73%
RotationFor	53,61%	54,57%	54,57%	53,06%
LogitBoost	53,88%	54,62%	54,48%	53,30%
BayesNet	53,64%	54,55%	54,45%	53,59%
NaiveBayes	53,64%	54,38%	54,43%	53,37%
HomeWins	48,66%	48,66%	48,66%	48,66%
Classifier/features-->	**3,4**	**5,6**	**1-6**	**7-10**
ClassViaReg	53,42%	54,40%	54,86%	53,13%
MultiCC	53,40%	54,38%	54,84%	52,92%
RotationFor	52,92%	54,00%	54,04%	53,56%
LogitBoost	53,35%	53,78%	54,00%	53,28%
BayesNet	53,28%	54,57%	54,38%	52,82%
NaiveBayes	53,23%	54,36%	54,43%	52,99%
HomeWins	48,66%	48,66%	48,66%	48,66%

A.3 Feature selection

The first part of the feature selection dealt with the same feature set used in the previous section, in which, one at a time, features were left out. Table 6 shows the performance of our chosen classifier when one of the features was left out.

Table 6. Classification performance with one feature left out

Feature left out	**Performance**
none	55,00%
1	54,93%
2	54,96%
3	54,93%
4	54,96%
5	54,74%
6	55,00%
7	55,05%
8	55,09%
9	55,00%
10	54,88%

B. DATA EVALUATION

In this appendix we describe the formulas behind the calculating of the probabilities used for the betting strategies. Also the results of the betting strategies are shown in full detail.

B.1 Calculating probabilities

The formulas the classifier comes up with for each of the outcomes are the following (the numbers between the parentheses indicate the features from the earlier given feature list):

Home win: 0.0042 * (1) - 0.0049 * (2) - 0.0041 * (3) + 0.0078 * (4) + 0.0842 * (5) - 0.0651 * (6) + 0.0033 * (9) - 0.2385 * (18) + 0.6477

Draw: 0.0022 * (2) - 0.0025 * (4) - 0.0052 * (9) - 0.0059 * (10) + 0.4009

Away win: -0.0029 * (1) + 0.0021 * (2) + 0.0038 * (3) - 0.0056 * (4) - 0.0904 * (5) + 0.0481 * (6) + 0.0083 * (10) + 0.1853 * (18) + 0.0585

After the formulas had been evaluated for all matches in the database, these numbers (which we will for convenience' sake call *a*, *b* and *c*) were normalized (into a_{norm}, b_{norm} and c_{norm}) as follows:

a_{norm} = a/(a+b+c)

b_{norm} = b/(a+b+c)

c_{norm} = c/(a+b+c)

B.2 Betting strategies

Table 7 shows the results of all the betting strategies that were evaluated.

Table 7. Performance of the betting strategies

Strategy	last ½ season	last season	all 7 seasons
I	4.41	5.44	-25.784
II	-11.13	-17.95	-119.716
III-0 or IV-1	-0.99	-4.86	12.486
III-0.02	-1.39	-2.76	26.099
III-0.04	1.25	4.82	41.133
III-0.06	4.09	6.42	45.177
III-0.08	1.15	-1.32	48.214
III-0.10	1.9	-0.32	40.144
IV-1.04	-2.92	-4.65	22.669
IV-1.08	-1.29	-1.31	25.203
IV-1.12	-2.41	-0.7	31.414
IV-1.16	3.68	4.51	56.504
IV-1.20	0.9	-1.57	40.82
IV-1.25	0.9	0.43	31.79