Рейтинговые системы для прогнозирования футбольных матчей с фиксированными коэффициентами

*Что такое рейтинговая система?*

Система рейтингов представляет собой количественную меру превосходства одной футбольной команды над соперником в матче. Такое превосходство определяется путем анализа и сравнения одного или нескольких аспектов прошлого выступления каждой из сторон. Рейтинговые системы различаются по способу расчета превосходства сторон, но в основном каждый метод рассчитывает разницу очков для предстоящего футбольного матча путем вычитания рейтинга очков гостевой команды из рейтинга очков домашней команды. Рейтинги очков домашней и гостевой команд определяются путем количественного анализа прошлых результатов с учетом различных аспектов силы команды. Самые простые из них используют либо очки лиги, либо положение в лиге, либо пропущенные и забитые голы, в то время как более сложные рейтинги могут быть основаны на подробной статистике матча, включая удары по воротам, угловые и, возможно, даже владение мячом, если такие данные доступны.

Во многих простых рейтинговых системах не учитывается качество соперника. Например, система рейтинга разницы забитых и пропущенных мячей в последних матчах просто рассматривает количество забитых и пропущенных мячей двумя командами за определенное количество матчей, предшествующих исследуемому поединку. Один гол всегда стоит одного гола, независимо от того, забит ли он на выезде "Манчестер Юнайтед" или дома "Бирмингем Сити". Power Ratings или Rateform, описанные Биллом Хантером в его книге  [*Football Fortunes*](http://www.gamblingbooks.co.uk/stores/footballdata/titleinfo.php?item=903)*, позволяют* решить эту проблему путем пропорционального соотношения стоимости каждого забитого (или пропущенного) гола и силы соперника, против которого он был забит (или пропущен).

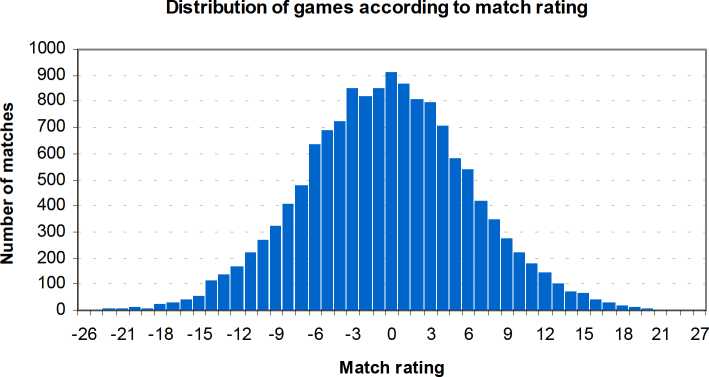
После расчета рейтинга матча следующим шагом будет оценка вероятности каждого результата, на основании которой можно делать прогнозы ставок. Рейтинг матча должен быть каким-то образом переведен в распределение вероятностей для трех возможных результатов футбольного матча: победа хозяев, ничья и победа гостей. Для этого необходимо проанализировать некоторые исторические данные, которые можно найти на [Football-Data.](http://www.football-data.co.uk/) Этот процесс описан для простой системы оценки превосходства по голам за последние годы.

*Система рейтинга превосходства целей*

Разница мячей - это один из показателей доминирования одной футбольной команды над другой в матче. Таким образом, система оценки превосходства по голам предполагает, что команды, которые забивают больше голов и пропускают меньше в течение нескольких матчей, с большей вероятностью выиграют свою следующую игру. Иными словами, их игровая форма с точки зрения забитых и пропущенных голов потенциально лучше, чем у команд с более низким рейтингом. Обычно под последней формой подразумеваются последние 4, 5 или 6 матчей. В данном случае выбраны последние 6 матчей.

Чтобы увидеть, как рассчитывается рейтинг превосходства по голам, рассмотрим следующий пример игры между "Тоттенхэмом" и "Лидсом" на "Уайт Харт Лейн". В последних 6 матчах "Тоттенхэм" забил 6 голов и пропустил 9. В то же время "Лидс" забил 8 раз и пропустил 11 голов. Рейтинг превосходства "Тоттенхэма" по голам за последние 6 матчей составляет -3; у "Лидса" он также равен -3. Рейтинг матча определяется как рейтинг хозяев минус рейтинг гостей, поэтому для данного матча он равен 0. Имея достаточно исторических данных, мы можем перевести этот рейтинг матча в распределение вероятности победы хозяев, ничьей и победы гостей, а затем определить справедливые коэффициенты на один или несколько из трех возможных результатов. Чтобы начать эту задачу, мы должны посмотреть, как часто матч с рейтингом превосходства по голам 0 заканчивается победой дома, ничьей и победой в гостях.

Используя данные о результатах английской Премьер-лиги и Дивизионов 1, 2 и 3 за сезоны с 1993/94 по 2000/01, были рассчитаны рейтинги превосходства по голам для последних 6 матчей, сыгранных каждой командой. Из 16 272 матчей, сыгранных за эти 8 лет, 14 002 матча были пригодны для расчета рейтинга, причем матчи, сыгранные в начале каждого сезона, явно не подходят для анализа последней формы. Количество матчей с каждым рейтингом показано на рисунке ниже. Рейтинги варьируются от -26 (очень слабая домашняя команда и сильная гостевая) до +27 (очень сильная домашняя команда и слабая гостевая). Очевидно, что игр с большим положительным или отрицательным рейтингом гораздо меньше. Из этих игр 46,2% завершились победой хозяев, 28,1% - ничьей и 25,7% - победой гостей.



Количество и процент побед дома, ничьих и побед на выезде для каждого рейтинга матча представлены в таблице на следующей странице. При рейтинге матча 0 мы можем предположить, что "Тоттенхэм" имеет примерно 46% шансов на победу, так как из 908 матчей, сыгранных с 1993 по 2001 год и имевших такой же рейтинг, почти 46% закончились победой хозяев. Соответствующий коэффициент на победу "Лидса" составляет 26%, а на ничью - 28%. Это практически совпадает с долгосрочным средним распределением результатов: 46,2%, 28,1% и 25,7% для побед дома, ничьих и побед на выезде соответственно для всей английской футбольной лиги, что неудивительно, поскольку нулевой рейтинг матча находится примерно на полпути между самым высоким и самым низким рейтингом матча.

Поскольку мы рассчитали распределение вероятностей результатов для каждого рейтинга матча, мы можем использовать табличные данные на следующей странице, чтобы предсказать наиболее вероятный результат любого матча, при условии, что в сезоне было сыграно не менее 6 игр, чтобы описать недавнюю форму. Согласно этим данным, матч с рейтингом +12, например, имеет 63% шансов закончиться победой хозяев, в то время как в другом матче с рейтингом -12 победа команды гостей ожидается с вероятностью 39%. В целом, чем выше рейтинг матча, тем выше вероятность победы хозяев. И наоборот, чем ниже рейтинг, тем больше шансов на победу в гостях. Изначально неочевидно, как рейтинг матча влияет на вероятность ничьей.

Рейтинг матчей по превосходству в голах и исторические проценты результатов

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| *Рейтинг матча* | *Количество домашних побед* | *Количество розыгрышей* | *Количество побед на выезде* | *% домашних побед* | *% домашних ничьих* | *% побед на выезде* |
| -26 | 0 | 1 | 1 | 0.0% | 50.0% | 50.0% |
| -23 | 0 | 0 | 2 | 0.0% | 0.0% | 100.0% |
| -22 | 0 | 0 | 3 | 0.0% | 0.0% | 100.0% |
| -21 | 0 | 2 | 4 | 0.0% | 33.3% | 66.7% |
| -20 | 2 | 2 | 7 | 18.2% | 18.2% | 63.6% |
| -19 | 1 | 1 | 3 | 20.0% | 20.0% | 60.0% |
| -18 | 5 | 7 | 9 | 23.8% | 33.3% | 42.9% |
| -17 | 7 | 9 | 12 | 25.0% | 32.1% | 42.9% |
| -16 | 6 | 14 | 21 | 14.6% | 34.1% | 51.2% |
| -15 | 25 | 12 | 19 | 44.6% | 21.4% | 33.9% |
| -14 | 32 | 30 | 51 | 28.3% | 26.5% | 45.1% |
| -13 | 43 | 38 | 58 | 30.9% | 27.3% | 41.7% |
| -12 | 51 | 50 | 64 | 30.9% | 30.3% | 38.8% |
| -11 | 75 | 54 | 91 | 34.1% | 24.5% | 41.4% |
| -10 | 84 | 94 | 91 | 31.2% | 34.9% | 33.8% |
| -9 | 123 | 91 | 112 | 37.7% | 27.9% | 34.4% |
| -8 | 171 | 113 | 124 | 41.9% | 27.7% | 30.4% |
| -7 | 190 | 121 | 170 | 39.5% | 25.2% | 35.3% |
| -6 | 242 | 202 | 191 | 38.1% | 31.8% | 30.1% |
| -5 | 279 | 212 | 197 | 40.6% | 30.8% | 28.6% |
| -4 | 293 | 219 | 215 | 40.3% | 30.1% | 29.6% |
| -3 | 374 | 246 | 229 | 44.1% | 29.0% | 27.0% |
| -2 | 372 | 233 | 214 | 45.4% | 28.4% | 26.1% |
| -1 | 375 | 251 | 222 | 44.2% | 29.6% | 26.2% |
| 0 | 414 | 259 | 235 | 45.6% | 28.5% | 25.9% |
| 1 | 412 | 243 | 212 | 47.5% | 28.0% | 24.5% |
| 2 | 401 | 220 | 189 | 49.5% | 27.2% | 23.3% |
| 3 | 395 | 224 | 175 | 49.7% | 28.2% | 22.0% |
| 4 | 391 | 177 | 137 | 55.5% | 25.1% | 19.4% |
| 5 | 297 | 180 | 102 | 51.3% | 31.1% | 17.6% |
| 6 | 260 | 146 | 131 | 48.4% | 27.2% | 24.4% |
| 7 | 236 | 98 | 83 | 56.6% | 23.5% | 19.9% |
| 8 | 197 | 94 | 56 | 56.8% | 27.1% | 16.1% |
| 9 | 158 | 86 | 32 | 57.2% | 31.2% | 11.6% |
| 10 | 125 | 57 | 42 | 55.8% | 25.4% | 18.8% |
| 11 | 113 | 34 | 33 | 62.8% | 18.9% | 18.3% |
| 12 | 90 | 30 | 22 | 63.4% | 21.1% | 15.5% |
| 13 | 61 | 23 | 17 | 60.4% | 22.8% | 16.8% |
| 14 | 48 | 15 | 11 | 64.9% | 20.3% | 14.9% |
| 15 | 38 | 21 | 8 | 56.7% | 31.3% | 11.9% |
| 16 | 30 | 9 | 2 | 73.2% | 22.0% | 4.9% |
| 17 | 20 | 8 | 2 | 66.7% | 26.7% | 6.7% |
| 18 | 15 | 1 | 1 | 88.2% | 5.9% | 5.9% |
| 19 | 8 | 4 | 1 | 61.5% | 30.8% | 7.7% |
| 20 | 5 | 1 | 0 | 83.3% | 16.7% | 0.0% |
| 21 | 1 | 0 | 0 | 100.0% | 0.0% | 0.0% |
| 22 | 1 | 0 | 1 | 50.0% | 0.0% | 50.0% |
| 23 | 1 | 0 | 0 | 100.0% | 0.0% | 0.0% |
| 27 | 1 | 0 | 0 | 100.0% | 0.0% | 0.0% |
| **Всего** | **6468** | **3932** | **3602** | **46.2%** | **28.1%** | **25.7%** |

Но как насчет рейтинга +15? Конечно, можно ожидать, что шансы на победу дома будут выше, чем при рейтинге +12, и все же только 57% матчей с рейтингом +15 завершились с таким результатом. Аналогично, только 34% матчей с рейтингом -15 завершились победой на выезде. Конечно, эти расхождения возникают потому, что связь между рейтингом матча и вероятностью результата по своей природе "шумная" и несовершенная. Такие расхождения становятся более очевидными для экстремальных рейтингов, для которых из-за ограниченного количества данных о матчах один или два результата оказывают гораздо большее влияние на распределение вероятности результата. Чтобы учесть это расхождение, нам необходимо стандартизировать нашу модель прогнозирования. Таким образом, мы можем предпринять практическую попытку определить справедливые коэффициенты для футбольного матча.

*Определение справедливых коэффициентов*

Первая задача состоит в том, чтобы рассмотреть каждый результат независимо и определить "наилучшую" связь с рейтингами матчей. Самый простой способ определить эту наилучшую взаимосвязь - построить рейтинги матчей и вероятности результатов в виде трех диаграмм рассеяния, по одной для каждого результата, как показано на следующей странице. Это можно сделать с помощью программы Microsoft Excel. Для каждой диаграммы рассеяния на точки данных накладывается линия наилучшего соответствия (с ее уравнением), представляющая то, что статистически считается наилучшей взаимосвязью между рейтингом матча и вероятностью результата.

Уравнение для каждой линии наилучшего соответствия также можно легко рассчитать с помощью Microsoft Excel, и оно показано для каждого рисунка. В каждом уравнении y, вероятность наступления определенного результата, является некоторой функцией x, рейтинга матча. Для домашних побед соответствующий рисунок на следующей странице ясно показывает, что наилучшей зависимостью является прямая линия. Значение R2, показанное для каждого из трех уравнений, является просто статистической мерой того, насколько точно реальные данные соответствуют наилучшим линиям. Идеальная зависимость, обозначаемая R2 = 1, означает, что линия и уравнение наилучшим образом описывают реальные данные. Следовательно, между рейтингом матча и вероятностью победы дома существует достаточно хорошая связь, где 86% вариаций реальных данных объясняются уравнением, которое лучше всего подходит к реальным данным. Для выездных побед и особенно ничьих связь слабее.

С помощью каждого уравнения мы можем легко определить ожидаемую вероятность победы дома, ничьей и победы на выезде для любого матча, в котором мы рассчитали рейтинг превосходства по голам. Например, для матча Тоттенхэм - Лидс, где рейтинг матча был равен 0, вероятность домашней победы можно определить, используя:

y = 1.56x + 46.47

где y = вероятность домашней победы, а x = рейтинг матча.

При рейтинге матча 0 вероятность домашней победы составит 46,47%. Это очень близко к проценту исторических матчей с рейтингом 0, закончившихся победой хозяев. Такое близкое совпадение возникает потому, что наилучшая зависимость обеспечивает особенно хорошее соответствие реальным данным (R2 = 0,8623). Прелесть этого уравнения в том, что нам больше не нужно помнить всю информацию, содержащуюся в таблице на предыдущей странице, чтобы рассчитать вероятность домашней победы для других игр с разными рейтингами. Например, в игре с рейтингом +10 вероятность домашней победы составляет 62,1%, а в игре с рейтингом -7 - 35,6%. Чтобы рассчитать эти проценты, достаточно ввести рейтинг матча в приведенное выше уравнение.

**Распределение домашних побед по рейтингу матча**

**% ничьих %побед в гостях %побед дома**

100%

♦ I ♦♦♦♦TT0%

-30-25-20-15-10-5051015202530

**Рейтинг матча**

**Распределение побед на выезде по рейтингу матча**

100%

90%

80%

10%

0%

-30-25-20-15-10-5051015202530

**М атч Рейтинг**

**Распределение ничьих по рейтингу матча**

100%

90% -

80% -

70% -

60% -

50% -

40% -

y = -0,03x2 - 0,29x + 29,48  
R2 = 0,3918

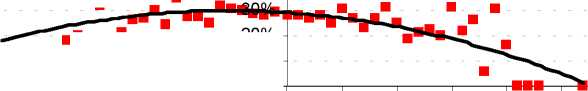
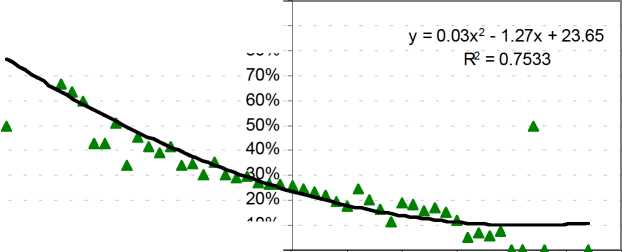
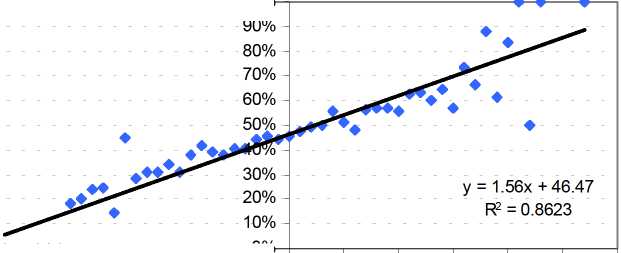
20% -

10% -

0%

-30-25-20-15-10-5051015202530

**М атч Рейтинг**



Оценив вероятность или ожидание победы дома, мы можем легко определить справедливые коэффициенты на победу дома, используя:

*100 разделить на вероятность победы дома*

Следовательно, справедливый коэффициент на победу "Тоттенхэма", согласно этой модели прогнозирования, составляет 100/46,47 или 2,15. Мы можем определить справедливые коэффициенты на выездную победу и ничью таким же образом, используя уравнения для выездных побед и ничьих (показанные с соответствующими цифрами на предыдущей странице), чтобы сначала рассчитать ожидание выездной победы и ничьей. Их точность, конечно, будет ниже, чем для домашних побед, так как уравнения с наилучшей подгонкой менее точно описывают реальные данные. Таким образом, справедливые коэффициенты на матч Тоттенхэм - Лидс составляют 2.15, 3.39 и 4.23 на победу хозяев, ничью и выездную победу соответственно. Справедливые коэффициенты для некоторых других рейтингов матчей представлены в таблице ниже.

Ожидания результата и их справедливые шансы для системы оценки превосходства цели

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| *Рейтинг матча* | *Ожидаемая победа на домашней арене* | *Ожидаемая продолжительность рисования* | *Ожидаемая победа на выезде* | *Всего* | *Справедливые домашние ставки* | *Справедливые коэффициенты на ничью* | *Справедливые коэффициенты на выезд* |
| -16 | 22% | 26% | 52% | 100% | 4.65 | 3.78 | 1.94 |
| -14 | 25% | 28% | 47% | 100% | 4.06 | 3.62 | 2.11 |
| -12 | 28% | 29% | 43% | 100% | 3.60 | 3.49 | 2.31 |
| -10 | 31% | 29% | 39% | 100% | 3.24 | 3.40 | 2.54 |
| -8 | 34% | 30% | 36% | 100% | 2.94 | 3.35 | 2.80 |
| -6 | 37% | 30% | 32% | 100% | 2.69 | 3.32 | 3.09 |
| -4 | 40% | 30% | 29% | 100% | 2.49 | 3.32 | 3.42 |
| -2 | 43% | 30% | 26% | 100% | 2.31 | 3.34 | 3.80 |
| 0 | 46% | 29% | 24% | 100% | 2.15 | 3.39 | 4.23 |
| 2 | 50% | 29% | 21% | 100% | 2.02 | 3.47 | 4.71 |
| 4 | 53% | 28% | 19% | 100% | 1.90 | 3.59 | 5.25 |
| 6 | 56% | 27% | 17% | 100% | 1.79 | 3.75 | 5.84 |
| 8 | 59% | 25% | 15% | 100% | 1.70 | 3.96 | 6.49 |
| 10 | 62% | 24% | 14% | 100% | 1.61 | 4.24 | 7.17 |
| 12 | 65% | 22% | 13% | 100% | 1.53 | 4.61 | 7.86 |
| 14 | 68% | 20% | 12% | 100% | 1.46 | 5.12 | 8.51 |
| 16 | 71% | 17% | 11% | 100% | 1.40 | 5.83 | 9.08 |

*Определение валуйных ставок*

После того как мы определили справедливые коэффициенты на матч, последний шаг будет самым простым. Сравнив наши справедливые коэффициенты с несправедливыми коэффициентами букмекера (с учетом его оверквотинга для получения прибыли), мы можем определить, является ли его ставка выгодной. Если коэффициенты букмекера превосходят наши справедливые коэффициенты, мы потенциально получаем преимущество, при условии, что наша система рейтингов дает точный прогноз или, что более уместно, точное ожидание результата. В таком случае букмекер недооценил вероятность результата и предлагает, по оценке нашего прогноза, более высокие коэффициенты, чем математически справедливые. В долгосрочной перспективе, при условии точности анализа рейтингов, необходимо получать прибыль. Различие между прогнозом результата и ожиданием результата очень важно и является центральным для принципов value betting. Недостаточно просто предсказать наиболее вероятный результат. Чтобы потенциально получить прибыль за серию ставок, мы должны всегда сравнивать свои справедливые коэффициенты с коэффициентами букмекера и рассматривать исход матча с точки зрения распределения вероятностей, а не просто как победу или поражение.

На игру между "Тоттенхэмом" и "Лидсом" букмекерская контора Interwetten предлагает коэффициент 2,2 на победу "Тоттенхэма". Поскольку эти коэффициенты выше, чем рассчитанный нами справедливый коэффициент 2,15, это была выгодная ставка. Игра закончилась со счетом 2:0 в пользу "Тоттенхэма", и ставка оказалась выигрышной. Однако одна выигрышная ставка не является показателем прибыльности системы. Чтобы убедиться в том, что у нас есть такая система, любую систему следует тестировать на более длительных сериях матчей, например, в течение всего сезона. Имея доступ к историческим результатам и коэффициентам ставок от  [Football-Data,](http://www.football-data.co.uk/) можно проверить это ретроспективно.

*Тестирование системы прогнозирования матчей*

Ниже приведены результаты воображаемого учета ставок с использованием системы рейтинга превосходства по голам для определения значимости домашних побед в матчах английской лиги, сыгранных в сезоне 2001/02. Всего 1746 игр получили рейтинг матча, причем первые 6 игр команды явно не подлежали оценке. Там, где была найдена выгодная ставка, была сделана единичная ставка (1 очко). Поскольку при прогнозировании матчей с очень низким или очень высоким рейтингом существует большая степень неопределенности (так как гораздо меньше матчей с таким рейтингом участвовало в определении уравнения для прогнозирования), можно предпочесть избегать ставок на такие матчи. И наоборот, прогнозы на матчи с рейтингом, близким к центральному значению 0, более надежно определяются уравнением наилучшего соответствия (см. верхний рисунок на странице 5). Таким образом, описывается анализ прибыли, противопоставляющий применение в ставках четырех различных диапазонов рейтингов матчей.

Анализ прибыли для модели прогнозирования рейтингов матчей с превосходством голов, только домашние победы

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Все рейтинги | Рейтинги от -10 до +10 | Рейтинги от -5 до +5 | Рейтинги от -2 до +2 | Ставки вслепую на все домашние победы |
| Количество ставок | 526 | 459 | 330 | 171 | 1746 |
| Прибыль от лучших коэффициентов | **+11.07** | **+16.32** | **+17.65** | **+17.32** | **-65.20** |
| Доходность от лучших шансов | **+2.10%** | **+3.56%** | **+5.35%** | **+10.13%** | **-3.73%** |

Принимая лучшие доступные коэффициенты (от 6 букмекеров), система рейтинга превосходства голов принесла бы прибыль над оборотом (или доходность) в течение футбольного сезона 2001/02 в Англии в размере 2,1% для всех рейтингов матчей. Этот показатель увеличивается по мере сужения диапазона рейтингов матчей к центральному рейтингу 0. Для матчей с рейтингом от -2 до +2 на каждый поставленный фунт стерлингов приходилось бы более 1,10 фунта стерлингов. Это сравнимо с доходностью -3,73%, если ставить на каждую домашнюю победу вслепую.

Аналогичный анализ прибыли можно провести для выездных побед и ничьих. Большая неопределенность в модели прогнозирования, как уже упоминалось ранее, сделает прогнозы менее надежными, а прибыль - более труднодостижимой.

Данное руководство по прогнозированию матчей составлено на основе переработанного материала из готовящейся к изданию книги Джо Бучдала "*Ставки на спорт с фиксированными коэффициентами: The Essential Guide, которая* будет издана High Stakes Ltd. и должна выйти в свет осенью-зимой 2003 года.