**КОМПЬЮТЕРНЫЕ ИССЛЕДОВАНИЯ
И МОДЕЛИРОВАНИЕ 2014 Т. 6 № 3 С. 455−460**

Ки&М

МОДЕЛИ ЭКОНОМИЧЕСКИХ И СОЦИАЛЬНЫХ СИСТЕМ

УДК: 519.67, 519.245

Об одном подходе к имитационному моделированию спортивной игры с непрерывным временем

Р. Б. Прядеин[1], М. Е. Степанцов[2,a]

[1] Национальный исследовательский университет «Высшая школа экономики»,
факультет бизнес-информатики, отделение прикладной математики и информатики
Россия, 123022, г. Москва, Большой Трёхсвятительский пер., 3
[2] Национальный исследовательский университет «Высшая школа экономики»,
факультет экономики, департамент математики
Россия, 119049, г. Москва, ул. Шаболовка, д. 26, корп. 5

E-mail: [a] mews@yandex.ru

*Получено 2 февраля 2014 г.,
после доработки 20 апреля 2014 г.*

Работа посвящена обсуждению методов статистического моделирования исходов спортивных соревнований вообще и спортивной игры с непрерывным временем в частности. Предложен основанный на имитационном моделировании хода такой игры подход к предсказанию результата игры, представляющий собой некоторый промежуточный вариант между чистым статистическим моделированием и агентным моделированием действий отдельных игроков, участвующих в матче. Приведен пример ретроспективного прогноза на основе предложенной модели.

Ключевые слова: математическое моделирование, имитационное моделирование, статистическое моделирование, спортивные соревнования

On a possible approach to a sport game with continuous time simulation

R. B. Priadein[1], M. Ye. Stepantsov[2]

[1] National Research University Higher School of Economics, Faculty of Business Informatics, Department of Applied Mathematics and Informaticss, 3 Bolshoi Trekhsvyatitelskiy, Moscow, 123022, Russia
[2] National Research University Higher School of Economics, Faculty of Economics, Department of Mathematics, 26 Shabolovka, building 5, Moscow, 119049, Russia

Abstract. — This paper is dedicated to discussing methods of statistical modeling the outcomes of sport events and, particularly, matches with continuous time. We propose a simulation-based approach to predicting the outcome of a match, somehow medium between pure statistical methods and agent simulation of individual players. An example of retrospective prediction is given.

Keywords: mathematical modeling, simulation, statistical modeling, sport events

Citation: *Computer Research and Modeling*, 2014, vol. 6, no. 3, pp. 455–460 (Russian).

Одна из причин того, что спортивные соревнования вообще и матчи в игровых видах спорта в частности вызывают у людей большой интерес, состоит в непредсказуемости исхода спортивной борьбы. Тем не менее в некоторых случаях важна как раз возможность с достаточной вероятностью предсказать результат состязания. Организаторам соревнований и телевизионных или онлайн-трансляций полезно предполагать, как сложится ход игры и какие команды смогут выйти в ту или иную стадию соревнований. Спортивным функционерам и букмекерам важно знать вероятности различных исходов матчей, чтобы оперативно отслеживать и предупреждать такое печальное явление как договорные матчи. Возможно, и тренерам команд могло бы быть полезно знать идеальный результат предстоящей встречи, чтобы каким-либо образом скорректировать игру своей команды, а по итогам реального матча сделать выводы о факторах, приведших к тому или иному исходу противостояния.

Конечно, при наличии такой потребности в предсказании результатов игр не вызывает удивления, что были разработаны многочисленные математические модели спортивных соревнований, позволяющие строить прогнозы. Однако здесь следует отметить, что такие модели в основном представлены двумя крайностями. С одной стороны это статистические модели, рассматривающие матч как одно событие с дискретным набором исходов и фиксированными вероятностями каждого из исходов [Reep, Benjamin, 1968]. Задача прогнозирования в этом случае состоит просто в получении оценок этих вероятностей на основе имеющейся статистики результатов игр [Dixon, Coles, 1997].

С другой стороны существует подход, основанный на имитационном моделировании спортивных качеств каждого из спортсменов, участвующих в игре. Ярким представителем такого подхода является игровая система FIFA [Electronic Arts]. Хотя данная модель и преследует прежде всего развлекательные цели, она и ее аналоги широко используются для прогнозирования результатов матчей. При применении такого подхода результаты моделирования подвержены сильному влиянию допущений, проведенных при моделировании игроков, поскольку не подлежит сомнению тот факт, что адекватная математическая модель отдельного человека, пусть даже и с крайне ограниченным списком существенных свойств, пока что не создана.

В данной работе предлагается рассмотреть предложение некоего среднего между двумя указанными крайностями подхода, состоящего в том, что матч представляется в виде марковской цепи событий [Кемени, Снелл, 1970] с дискретным или непрерывным временем (в зависимости от того, является ли матч непрерывным по времени или состоит из отдельных розыгрышей очков). Данная работа содержит предварительное предложение модели, требующей дальнейшей апробации и доработки по результатам такой апробации. При осуществлении предлагаемого подхода моделируется не матч в целом, но и не отдельные действия отдельных игроков, а изменение счета, то есть перехода системы из одного состояния в другое. Принимая во внимание условность любой классификации, тем не менее можно сформулировать этот подход как использование не системного и не агентного, а дискретно-событийного моделирования. В качестве вида спорта, иллюстрирующего марковскую цепь с непрерывным временем, будут рассмотрены соответственно футбол и большой теннис.

Построим вначале статистическую модель ожидаемого числа голов в футбольном матче. Рассмотрим количества мячей, забитых в каждом из матчей в ворота каждой команды, как случайные величины. Для каждой из команд, участвующих в двухкруговом турнире (то есть турнире, в котором каждая команда играет с каждой два матча — дома и в гостях), введем следующие показатели.

Сила атаки дома:

$$AH_i = \frac{\overline{SH_i}}{\overline{\overline{SH}}}. \tag{1}$$

Сила атаки в гостях:

$$AA_i = \frac{\overline{SA_i}}{\overline{\overline{SA}}}. \tag{2}$$

Сила защиты дома:

$$DH_i = \frac{\overline{SA}}{\overline{CH_i}}. \tag{3}$$

Сила защиты в гостях:

$$DA_i = \frac{\overline{SH}}{\overline{CA_i}}. \tag{4}$$

Здесь для команды i $\overline{SH_i}$ — математическое ожидание числа мячей, забитых дома, $\overline{SA_i}$ — математическое ожидание числа мячей, забитых в гостях, $\overline{CH_i}$ — математическое ожидание числа мячей, пропущенных дома, $\overline{CA_i}$ — математическое ожидание числа мячей, пропущенных в гостях, $\overline{SH}$ — математическое ожидание числа мячей, забитых дома (и, соответственно, пропущенных в гостях) в каждом матче лиги, $\overline{SA}$ — математическое ожидание числа мячей, забитых в гостях (и, соответственно, пропущенных дома) в каждом матче лиги.

После этого мы принимаем в качестве математического ожидания числа мячей, забитых хозяевами в конкретном матче между командой-хозяином I и командой-гостем j:

$$S_{ij} = \frac{AH_i}{DA_j}\overline{SH}. \tag{5}$$

Математическое ожидание числа пропущенных хозяевами мячей получаем аналогично:

$$C_{ij} = \frac{AA_j}{DH_i}\overline{SA}. \tag{6}$$

Подставив (1)–(4) в формулы (5)–(6) можно сократить по одному множителю в каждом случае, однако и технически, и содержательно удобнее работать именно с представленными выражениями.

В этой модели в качестве оценок всех математических ожиданий, отмеченных в (1)–(6) чертой сверху, используем средние значения по рассматриваемой футбольной лиге. Вопрос о том, за какой промежуток следует усреднять эти показатели, связан с выходящим за рамки данной работы обсуждением предположения о неизменности среднего уровня игры футбольной команды в течение определенного периода времени.

Рассмотрим пример построения имитационной схемы на основе такой модели футбольного матча на примере Английской футбольной премьер-лиги для матча команд «Ньюкасл» и «Тоттенхем», который открывал сезон 2012–2013. Статистические оценки показателей получим по итогам предыдущего (2011–2012) сезона [Футбол онлайн].

$$\overline{SH} = 1.589,$$
$$\overline{SA} = 1.216,$$
$$\overline{SH_1} = 1.526,$$
$$\overline{CA_2} = 1.263.$$

Получаем оценку математического ожидания числа мячей, забитых «Ньюкаслом»:

$$AH_1 = 0.960,$$
$$DA_2 = 1.258,$$
$$S_{12} = 1.213.$$

Аналогично находим математическое ожидание числа мячей, пропущенных в этом мачте «Ньюкаслом» (то есть забитых «Тоттенхемом»):

$$C_{12} = 1.046.$$

Далее обычно используется рассмотрение процесса забивания мячей как пуассоновского, в котором полученные нами оценки используются как математические ожидания пуассоновского распределения числа забитых мячей.

Однако обосновано ли такое предположение? Проверим данные по16 тысячам матчей ведущих Европейских чемпионатов (Испания, Германия, Франция, Англия, Россия), сыгранных с 2005 года на предмет распределения числа забитых мячей по Пуассону, используя критерий χ^2. Оказывается, что уровень значимости достигает здесь 0.1.

Таблица. 1. Реальная и пуассоновская частоты числа забитых мячей.

Число мячей	Реальная частота	Пуассоновская модель
0	0.0840	0.0764
1	0.1887	0.1964
2	0.2515	0.2526
3	0.2137	0.2166
4	0.1392	0.1393
5	0.0723	0.0717
6	0.0321	0.0307
7	0.0117	0.0113
8	0.0048	0.0036
>8	0.0021	0.0014

Если рассмотреть использовавшиеся для расчета критерия χ^2 частоты, становится ясно, что основные различия существуют для числа забитых мячей 0 (превышение над пуассоновской частотой) и 1 (обратная ситуация).

Следует отметить, что для различных чемпионатов уровень значимости получается существенно различным. Так, для чемпионата Франции он ниже 0.05, а для Английской футбольной премьер-лиги, напротив, превышает 0.1. Но во всех случаях различия частот характерны именно для значений 0 и 1.

Можно предположить, что подобная ситуация связана с тем, что при нулевом счете команды зачастую осторожничают, в то время как гол, забитый одной из команд, очень сильно мотивирует вторую команду к активным действиям и часто становится причиной второго гола (который может быть забит и в ворота второй команды, забывшей об обороне).

С другой стороны, статистическое исследование показывает, что нет значимой корреляции между количеством мячей, забитых хозяевами и гостями.

В связи с этим можно было бы предложить находить вероятности того или иного результата игры между двумя конкретными командами по формуле:

$$p(X,Y) = \text{Poisson}(x,\lambda_X)\text{Poisson}(y,\lambda_Y), \tag{7}$$

где $\text{Poisson}(x,\lambda_X)$ — вероятность принятия значения *x* случайной величиной *X*, подчиняющейся распределению Пуассона с параметром λ_X, $\text{Poisson}(y,\lambda_y)$ — вероятность принятия значения *y* случайной величиной *Y*, подчиняющейся распределению Пуассона с параметром λ_y.

В связи с вышеизложенным при разыгрывании результатов матча в рамках имитационного моделирования предлагается использовать исправленное распределение Пуассона с вероятностями счета X:Y, задаваемыми формулой

$$p^*(X,Y) = k(x,y,\lambda_{X,}\lambda_Y)\ \text{Poisson}(x,\lambda_X)\ \text{Poisson}(y,\lambda_Y), \tag{8}$$

где $k\left(x, y, \lambda_X, \lambda_Y\right)$ определяется следующим образом:

$$k\left(x, y, \lambda_X, \lambda_Y\right)=\begin{cases} 1+\alpha\lambda_X\lambda_y \text{ при } x=0, \quad y=0; \\ 1-\alpha\lambda_X \text{ при } x=0, \quad y=1; \\ 1-\alpha\lambda_y \text{ при } x=1, \quad y=0; \\ 1-\alpha \text{ при } x=1, \quad y=1; \\ 1 \text{ иначе,} \end{cases}$$

после чего для получения собственно вероятностей значения $p*$ нормируются на 1.

Коэффициент α зависит от рассматриваемого турнира и может быть подобран методом наименьших квадратов (минимизируется расстояние между векторами частот реального и исправленного распределения числа забитых мячей. В случае Английской футбольной премьер-лиги этот показатель равен 0.1.

После построения такого распределения подставляем в него значения среднего числа забитых мячей из формул (5)–(6) и в зависимости от задачи либо находим распределение вероятностей разных исходов моделируемого матча, либо разыгрываем этот исход как двумерную случайную величину в рамках имитационной схемы.

Например, в случае матча «Ньюкасл»–«Тотенхем» такое моделирование дало результат 2:1, соответствующий реальному исходу матча. Это, однако, можно объяснить только удачей, поскольку наиболее вероятным счетом в данном противостоянии был 1:1. Этот факт подтверждается и повторным имитационным моделированием матча. Как и следовало ожидать, модель дает 1:1 в качестве наиболее вероятного результата матча.

Однако в приведенной выше модели пока совершенно не учитывается тот факт, что футбольный матч имеет некоторую протяженность по времени и события в нем, в том числе голы, забитые *одной командой*, могут не являться независимыми. Напротив, есть все основания полагать, что успех команды (забитый мяч) в разных ситуациях может как повысить, так и понизить вероятности забить следующие мячи.

В соответствии с этим предлагается применять изложенный выше подход не ко всему матчу, а к отдельным отрезкам времени, на которые он может быть разбит. Для первого из таких отрезков модель (1)–(6) с поправкой (8) сохраняется в неизменном виде. Для каждого из следующих вычисляются условные средние значения числа забитых мячей, найденные при условии, что в течение предыдущих отрезков времени сложился тот или иной текущий результат матча и используется та же модель, но уже с условными значениями в качестве пуассоновских констант. Следует отметить, что такой подход требует обработки большого количества статистических данных, поскольку для каждого значения счета матча после *i*-го отрезка времени нужно собрать значимую статистику мячей, забитых в следующих отрезках. В рамках данной работы поэтому рассматривалось только деление матча на два тайма.

Но даже в этом случае собранная по матчам Английской футбольной премьер-лиги статистика представляет собой достаточно объемный массив данных, в связи с чем приведем только некоторую показательную ее часть.

В таблице 2 приведены условные вероятности наиболее часто встречающихся исходов второго тайма при известных результатах первого (остальные ситуации в использованном массиве данных встречались недостаточно часто, чтобы по ним можно было оценить соответствующие условные вероятности). Данные таблицы подтверждают, например, что ничейный счет первого тайма снижает активность команд во втором тайме.

Рассмотрим теперь матч «Ньюкасл»–«Тотенхем» при помощи этого подхода. Здесь, найдя для первого тайма величины $\overline{SH}=0.799$ и $\overline{SA}=0.608$ и сохраняя прежние значения силы атаки и защиты команд, путем моделирования получаем итог первого тайма 0:0. Затем, разыграв аналогично второй тайм, имеем в качестве его итога (и окончательного счета матча) 1:1. Результат этот, как кажется, не вполне соответствует реальному событию.

Таблица 2. Условные вероятности некоторых значений счета по итогам второго тайма в зависимости от счета по итогам первого

1-й тайм / 2-й тайм	0–0	1–0	0–1	1–1
0–0	0.271	0.229	0.213	0.239
1–0	0.195	0.184	0.194	0.212
0–1	0.128	0.155	0.168	0.142
1–1	0.118	0.116	0.108	0.109
2–0	0.087	0.084	0.083	0.072
0–2	0.045	0.049	0.059	0.054
2–1	0.043	0.049	0.049	0.052
1–2	0.031	0.042	0.040	0.031
2–2	0.010	0.015	0.019	0.010

Однако взглянем внимательнее на протокол матча. После первого тайма счет действительно был 0:0. Во втором тайме счет 1:1 установился и сохранялся до 80-й минуты, когда в ворота «Тотенхема» был назначен пенальти, успешно реализованный игроком «Ньюкасла». По итогам нашего моделирования можно сделать однозначный вывод, что этот назначенный одиннадцатиметровый удар оказал решающее влияние на результат матча.

Рассмотренная модель предлагается авторами в качестве предварительного варианта, требующего апробации, дополнительного анализа адекватности модели и возможной дальнейшей доработки по результатам такого анализа. В частности, такой доработкой может быть ее развитие в направлении большей детализации при разбиении матча на отрезки времени, что ставит вопрос о принципиально новом описании таких явлений, как, например, мотивация команды, поскольку при применении прежнего подхода — статистической обработки отдельных отрезков матча — объем необходимых сырых данных растет с увеличением числа отрезков разбиения экспоненциально.

Предполагается также рассмотреть динамику ряда параметров, считавшихся в первоначальном варианте модели по умолчанию неизменными, в том числе: усталость команды, наличие или отсутствие мотивации, настрой на командные действия и т. п. При этом статистика забитых мячей перестанет быть пуассоновской, что, как видно из формулы (8), имеет место уже и в предлагаемой здесь модели.

Проблемным является вопрос моделирования игроков, ранее не принимавших участия в турнире (или принципиально изменивших свои характеристики после предыдущих розыгрышей). В рамках данного подхода решение этой проблемы не представляется возможным, однако перспективной является идея сравнения среднего уровня силы участников рассматриваемых соревнований и соревнований, в которых новый игрок ранее принимал участие (модели, подобные рассматриваемой, обычно востребованы в применении к соревнованиям высшего уровня, участники которых до этого обычно участвуют в других турнирах).

Также предполагается обратить внимание на другие виды спорта, в частности такие, в которых течение матча принципиально дискретно и представляет собой конечное множество отдельных розыгрышей очков.

Список литературы

Кемени Дж., Снелл Дж. Конечные цепи Маркова. М.: Наука, 1970. — 272 с. с илл.

Чемпионаты, турнирные таблицы, результаты матчей // Футбол онлайн, 2013. URL: http://www.ftables.ru (дата обращения 11.03.2013).

Dixon M. J., Coles S. G. Modelling association football scores and inefficiencies in football betting market // Applied Statistics, 1997. — 46. — P. 265–280.

FIFA // Electronic Arts. URL: http://www.ea.com/fifa/ (дата обращения 01.11.2013).

Reep C., Benjamin B. Skill and chance in association football // Journal of the Royal Statistical Society, Series A, 1968. —131. — P. 581–585.