БУДІВЕЛЬНИ МЕХАНІКА ТА СИСТЕМНИЙ АНАЛІЗ

УДК 519.68; 681.513.7; 612.8.001.57; 007.51/.52

Гавриленко О.В., канд. фіз-мат. наук, **Дідківська В.А.**

ТЕХНОЛОГІЯ ПРОГНОЗУВАННЯ РЕЗУЛЬТАТІВ СПОРТИВНИХ ЗМАГАНЬ

Анотація. В даній роботі розглядається метод прогнозування результатів тенісних турнірів шляхом використання ланцюгів Маркова. У ході роботи було розглянуто стандартну реалізацію ланцюга Маркова за умови подавання гравцем і запропоновано модифікацію методу, у якій обрахунок ймовірності виграшу враховує ймовірність виграшу очка гравцем при прийомі м'яча.
Об'єкт дослідження – прогнозування результатів тенісних турнірів за допомогою ланцюгів Маркова.
Мета роботи – покращення способу перевірки даних та впровадження модифікації у алгоритм побудови ланцюгів Маркова.
Прогнозування результатів спортивних змагань являється актуальною дослідницькою задачею. Розробка методів та способів математичного моделювання дає змогу з достатньо високою точністю оцінити, яким буде результат змагання суперників. Одним із таких видів спорту є теніс – за протистоянням спортсменів щороку спостерігають мільйони шанувальників із усіх куточків світу. Із запалом дивляться такі змагання витривалості й вчені – комерційний та науковий інтерес зумовили сплеск досліджень у сфері прогнозування результатів тенісних турнірів. Більшість сучасних моделей для прогнозування тенісу використовують ієрархічні стохастичні вирази на основі ланцюгів Маркова. На даному етапі розвитку алгоритмів прогнозування результатів тенісних турнірів ланцюги Маркова являються менш комплексними у порівнянні з методами машинного навчання, за допомогою яких можна враховувати великий спектр параметрів матчу та історичну статистику. Однак, прогнозування на основі ланцюгів Маркова та модифікації даного методу можуть бути використані для оцінки результатів роботи методів прогнозування тенісних турнірів, заснованих на машинному навчанні. Перевірка правильності результатів новостворених методів являється важливою складовою процесу

дослідження, а, отже, повинні бути розглянуті ієрархічні стохастичні вирази на основі ланцюгів Маркова та запропоновані модифікації для оптимізації їх роботи.
Розглянуті методи прогнозування результатів тенісних турнірів допомагають оцінити результати навчання нейронних мереж алгоритмам прогнозування. Поліпшення даних алгоритмів підвищує точність прогнозування та достовірність отриманих результатів.
Ключові слова: прогнозування спортивних змагань, теніс, ланцюги маркова, теорія ймовірності, ймовірність виграшу

UDC 519.68; 681.513.7; 612.8.001.57; 007.51/.52

Havrilenko O., Cand. Phys.-Math. Sci. (Ph.D.), **Didkivska V.**

TECHNOLOGY FOR FORECASTING THE RESULTS OF SPORTS COMPETITIONS

Abstract. In this paper, the method of forecasting the results of tennis tournaments using the Markov chains is considered. In the course of the work, the standard implementation of the Markov chain was considered when the player serves and suggested a modification of the method, in which the calculation of the probability of winning takes into account the probability of winning a point by the player recepts.
The object of the study is to predict the results of tennis tournaments using Markov chains.
The goal of the work is to improve the way data is verified and to introduce a modification into the algorithm for constructing Markov chains.
Forecasting the results of sports competitions is a modern task for researchers. The development of methods and methods of mathematical modeling makes it possible to estimate with a high enough accuracy what will be the result of competition of rivals. One of these sports is tennis – for the confrontation of athletes annually observed by millions of fans from all over the world. They watch with such fervor such endurance competitions and scientists – commercial and scientific interest – caused a surge in research in the field of predicting the results of tennis tournaments. Most modern models for predicting tennis use hierarchical stochastic expressions based on Markov chains. At this stage of the development of algorithms for predicting the results of tennis tournaments, the Markov chains are less complex in comparison with the methods of machine learning, with the help of which one can take into account a large

range of match parameters and historical statistics. However, predictions based on Markov chains and modification of this method can be used to evaluate the results of the methods of forecasting tennis tournaments based on machine learning. Verification of the correctness of the results of the new methods is an important component of the research process, and therefore, hierarchical stochastic expressions based on Markov chains should be considered and modifications proposed to optimize their work.
The methods of forecasting the results of tennis tournaments are considered to help evaluate the results of training neural networks for prediction algorithms. The improvement of these algorithms increases the accuracy of forecasting and the reliability of the results obtained.
Keywords: forecasting of sports competitions, tennis, markova chain, probability theory, probability of winning

УДК 519.68; 681.513.7; 612.8.001.57; 007.51/.52

Гавриленко Е.В., канд. физ-мат. наук, **Дедковская В.А.**

ТЕХНОЛОГИЯ ПРОГНОЗИРОВАНИЯ РЕЗУЛЬТАТОВ СПОРТИВНЫХ СОРЕВНОВАНИЙ

Аннотация. В данной работе рассматривается метод прогнозирования результатов теннисных турниров путем использования цепей Маркова. В ходе работы были рассмотрены стандартную реализацию цепи Маркова при подачи игроком и предложено модификацию метода, в которой расчет вероятности выигрыша учитывает вероятность выигрыша очка игроком при приеме мяча.
Объект исследования - прогнозирование результатов теннисных турниров с помощью цепей Маркова.
Цель работы - улучшение способа проверки данных и внедрение модификации в алгоритм построения цепей Маркова.
Прогнозирование результатов спортивных соревнований является актуальной исследовательской задачей. Разработка методов и способов математического моделирования позволяет с достаточно высокой точностью оценить, каким будет результат соревнования соперников. Одним из таких видов спорта является теннис – за противостоянием спортсменов ежегодно наблюдают миллионы поклонников со всех уголков мира. С жаром смотрят такие соревнования выносливости и ученые - коммерческий и научный интерес обусловили всплеск исследований в области прогнозирования результатов теннисных турниров.

Большинство современных моделей для прогнозирования тенниса используют иерархические стохастические выражения на основе цепей Маркова. На данном этапе развития алгоритмов прогнозирования результатов теннисных турниров цепи Маркова являются менее комплексными по сравнению с методами машинного обучения, с помощью которых можно учитывать большой спектр параметров матча и историческую статистику. Однако, прогнозирования на основе цепей Маркова и модификации данного метода могут быть использованы для оценки результатов работы методов прогнозирования теннисных турниров, основанных на машинном обучении. Проверка правильности результатов вновь методов является важной составляющей процесса исследования, а, следовательно, должны быть рассмотрены иерархические стохастические выражения на основе цепей Маркова и предложены модификации для оптимизации их работы.

Рассмотрены методы прогнозирования результатов теннисных турниров помогают оценить результаты обучения нейронных сетей алгоритмам прогнозирования. Улучшение данных алгоритмов повышает точность прогнозирования и достоверность полученных результатов.

Ключевые слова: прогнозирование спортивных соревнований, теннис, цепи маркова, теория вероятности, вероятность выигрыша.

Вступ

Прогнозування результатів спортивних змагань являється актуальною дослідницькою задачею. Розробка методів та способів математичного моделювання дає змогу з достатньо високою точністю оцінити, яким буде результат змагання суперників. Одним із таких видів спорту є теніс. За протистоянням спортсменів щороку спостерігають мільйони шанувальників із усіх куточків світу. Із запалом дивляться такі змагання витривалості й вчені. Комерційний та науковий інтерес зумовили сплеск досліджень у сфері прогнозування результатів тенісних матчів зокрема і турнірів взагалі. Більшість сучасних моделей для прогнозування тенісу використовують ієрархічні стохастичні вирази на основі ланцюгів Маркова. На даному етапі розвитку алгоритмів прогнозування результатів тенісних турнірів ланцюги Маркова являються менш комплексними у порівнянні з методами машинного навчання, за допомогою яких можна враховувати великий спектр параметрів матчу та історичну статистику. Однак, прогнозування на основі ланцюгів Маркова та модифікації даного методу можуть бути використані для оцінки результатів роботи методів прогнозування тенісних турнірів, заснованих на машинному навчанні. Перевірка правильності результатів новостворених методів являється важливою складовою процесу дослідження, а, отже, повинні бути розглянуті ієрархічні стохастичні вирази на основі ланцюгів Маркова та запропоновані модифікації для оптимізації їх роботи.

Постановка задачі

Є тенісний матч проти одного суперника (сингл) – одиночний тенісний матч. Необхідно показати спосіб моделювання гейму даного матчу за допомогою ланцюгів Маркова, запропонувати покращення з точки зору підвищення точності прогнозу результатів.

Ланцюги Маркова у прогнозуванні тенісного турніру

Ланцюг Маркова – це випадковий процес, що задовольняє властивість Маркова і який приймає скінченну чи зліченну кількість значень (станів). Випадкова послідовність подій називається ланцюгом Маркова, якщо кожен перехід з одного стану в інший не залежить від того, коли і як система прийшла у поточний стан. Початковий стан може бути заданим заздалегідь або бути випадковим. Необхідно звернути увагу на те, що наступний стан залежить лише від поточного стану системи, усі попередні стани не враховуються – основуючись на цьому твердженні, можна побудувати ланцюг Маркова для тенісного матчу.

Імовірність виграшу очка у професійному тенісі являється ключовою у більшості способів прогнозування. Дана величина забезпечує обчислювальну базу, на якій засновується розрахунок ймовірності виграшу матчу. До наукової роботи, яка була написана Ф. Клаассеном та Й. Магнусом [1] вважалося, що очки матчу є незалежними й однаково розподіленими, тобто ймовірність виграшу очка константна протягом матчу. Вони ж, в свою чергу, перевірили дане припущення та спростували його, довівши, що очки в тенісі розподіляються незалежно та не однаково. Однак, відхилення від незалежних й однаково розподілених величин настільки незначне, що їх можна використовувати на практиці, так як незалежні й однаково розподілені величини дають зручні усереднені значення.

Сформулюємо математичну постановку задачі. Є два гравці тенісного матчу, перший гравець – A, другий гравець – B. Припустимо, що p – ймовірність того, що гравець A виграє очко, якщо подаватиме. Відповідно, припустимо, що q – ймовірність того, що гравець В виграє очко, якщо подаватиме. Нехай ймовірність виграшу очка незалежно та однаково розподілена.

Така постановка дає нам можливість побудувати ланцюг Маркова, за допомогою якого буде описано ймовірність виграшу гравця в геймі.

У тенісі гейм — відрізок сету, в якому право подачі належить виключно одному із супротивників, який повинен чергувати її, подаючи в ліву й праву половину корту. Стандартний гейм має наступний рахунок, рахунок гравця, що подає, називається першим:

– немає очок — «нуль»;
– перше очко — «п’ятнадцять»;
– друге очко — «тридцять»;
– третє очко — «сорок»;
– четверте очко — «гейм».

За винятком випадку, коли обоє гравців (пари) виграли по три очки, тоді рахунок оголошується «Рівне». Після «Рівне» рахунок стає «Більше» убік того гравця (пари), що виграв наступне очко. Якщо той же гравець виграє і наступне очко, цей гравець виграє «Гейм», якщо ж наступне очко виграє суперник, рахунок знову стає «Рівне». Гравець (пари) повинен виграти два очка поспіль після рахунку «Рівне», щоб виграти «Гейм».

Для перемоги у сеті потрібно виграти принаймні 6 геймів і принаймні на два більше, ніж супротивник. При рахунку 6:6 за геймами у більшості турнірів грається тай-брейк.

Вважатимемо рахунок простором станів, тоді переходи між станами – це ймовірності виграшу p першого гравця А або ймовірності виграшу q другого гравця В. Усі переходи, які означають очко, вигране гравцем А, мають ту ж ймовірність p, а всі переходи, які означають програне очко, мають ймовірність 1-p. Візуалізацію ланцюга Маркова за умови, що подає перший гравець А, наведено на рис.1 [2].

Виходячи з ідеї моделювання тенісних матчів з ланцюгами Маркова, Барнетт і Кларк [3] та О’Малі [4] розробили ієрархічні вирази для ймовівності того, що конкретний гравець виграє весь тенісний матч.

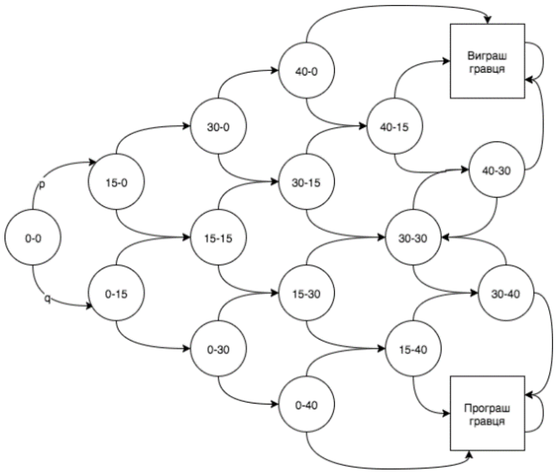

Рисунок 1 – Марківська модель гейму в матчі за умови, що подає гравець А

Барнетт і Кларк виражають ймовірність того, що гравець А виграв гру при своїй подачі $P_{гейм}$, використовуючи наступне рекурсивне визначення:

$$P_{гейм}(x,y) = p \cdot P_{гейм}(x+1,y) + (1-p)P_{гейм}(x,y+1). \qquad (1)$$

Граничні значення записуються як:

$$P_{гейм}(x,y) = 1, \text{ якщо } x = 4, \; x - y \geq 2. \qquad (2)$$

$$P_{гейм}(x,y) = 0, \text{ якщо } y = 4, \; y - x \geq 2. \qquad (3)$$

$$P_{гейм}(x,y) = \frac{p^2}{p^2 + (1-p)^2}, \text{ якщо } x = 4, \; x - y \geq 2. \qquad (4)$$

У наведеному вище, p – ймовірність того, що гравець А виграє очко при подачі, а x та y представляють кількість очок, виграних гравцями А та В, відповідно. Цей вираз явно відповідає ланцюжку Маркова.

Оцінка ймовірності виграшу при подачі

З огляду на ймовірність того, що обидва гравці виграють очко при їх подачі, ми можемо використовувати ієрархічні вирази, отримані Барнеттом і Кларком, щоб визначити переможця. Залишається питання, як оцінити ці ймовірності виграшу очка при подачі для ще не зіграних матчів. Барнетт і Кларк [3] дають ефективний метод для оцінки цих імовірностей з історичної статистики гравців:

$$f_i = a_i \cdot b_i + (1-a_i) \cdot c_i \qquad (5)$$

$$g_i = a_{av} \cdot d_i + (1-a_{av}) \cdot e_i, \qquad (6)$$

де f_i – відсоток очок, виграних при подачі гравцем i;

g_i – відсоток очок, виграних при прийомі м’яча гравцем i;

a_i – відсоток перших подач гравця i;

a_{av} – средний відсоток перших подач для всіх гравців;

b_i – відсоток виграшу при першій подачі гравця i;

c_i – відсоток виграшу при другій подачі гравця i;

d_i – відсоток виграшу при прийомі першої подачі гравцем i;

e_i – відсоток виграшу при прийомі другої подачі гравцем i.

Отже, для матчу між гравцями A і B ми можемо оцінити ймовірності виграшу очка при подачі гравцями A і B відповідно як f_{AB} і f_{BA}, використовуючи наступне рівняння:

$$f_{AB} = f_t + (f_i + f_{av}) - (g_i - g_{av}), \qquad (7)$$

де f_t – відсоток очок, виграних при подачі гравцем i;

f_{av} – відсоток очок, виграних при прийомі м'яча гравцем i;

g_{av} – відсоток перших подач гравця i.

Модифікація методу ланцюга Маркова до вирішення задачі прогнозування тенісних матчів може покращити отримувані результати та підвищити точність прогнозу. Кноттенбельт [5] уточнив моделі Барнета, використавши для обчислення ймовірності виграшу очка при подачі лише матчі зі спільними суперниками гравців, замість усіх суперників. Цей підхід дозволяє знизити похибку, що виникає через те, що гравці в минулому зустрічалися з суперниками різного рівня.

Мадурська [6] далі розширила модель загального суперника Кноттенбельта, використавши різні ймовірності виграшу очка при подачі для різних сетів. Таким чином, автор відмовилася від допущення IID і її модель відображає накопичення фізичної втоми у гравця по ходу матчу.

Моделювання тенісного гейму враховуючи статистичні дані гравця

В даній роботі, з метою підвищення точності прогнозу, наведено ще один варіант модифікації методу ланцюга Маркова.

Враховуючи те, що на даний момент існує вільний доступ до різноманітних статистичних даних тенісистів, то на початку матчу в якості ймовірності виграшу можем обрати відповідні дані з попередніх матчів або тренувань гравця (відсоток подачі (прийому), виграшів з подачі (прийому), тощо) [7]. Згідно з цим можна побудувати 2 дерева – для подачі та прийому гравцем м’яча. Окрім того, постійне оновлення статистичних даних дає змогу у кожному наступному геймі враховувати результати попереднього. Також слід зазначити, що таким чином можна враховувати статистику ейсів, відсотки попадання 1-ї та 2-ї подач та подвійних помилок, що робить схеми Маркова максимально докладними.

В якості прикладу проведемо модифікацію наведеного раніше способу побудови ланцюга Маркова для тенісного матчу між гравцем А та гравцем В [8].

Знайдемо умовні ймовірності отримання очок кожним із гравців та ймовірність виграшу гейму гравцем В за умови подачі та прийому, побудувавши таблиці умовних ймовірностей для обох віпадків та візуалізуємо їх марківськими моделями гейму.

Візьмемо за основу гейм і вважатимемо рахунок простором станів. Побудуємо модель для гравця В за умови, що не він подавав (прийняв) – оберемо деяке р, що означатиме ймовірність виграшу очка цим гравцем за умови, що він подає (приймає). Ймовірність виграшу очка суперником - гравцем А позначимо як q=l-p.

Для нашого прикладу, скориставшись статистичними даними матчу, встановлюємо, що ймовірність виграшу очка при подачі гравця В – це p=0,77, тоді ймовірність виграшу очка гравцем А – це q=1-p=0,23.

При прийомі гравця В – це p=0,41, тоді ймовірність виграшу очка гравця А – це q=1-p=0,59.

Побудуємо марківський процес із дискретним часом, кінцевим станом та матрицею переходів М і початковим станом I_0. Множина станів цього ланцюга являється поглинаючою (або замкненою), так як включає в себе 2 стани, з яких не може бути переходу в наступний стан – це стан виграшу гравця А та стан виграшу гравця В. Оскільки ми модифікуємо ситуацію гри гейму, то матриця М набуває вигляду:

$$M = \begin{bmatrix} m_{1,1} \ldots m_{1,17} \\ \ldots\ldots\ldots\ldots \\ m_{17,1} \ldots m_{17,17} \end{bmatrix}. \tag{8}$$

Це зумовлено тим, що при моделюванні марківського ланцюга для гейму гри може бути всього 17 станів системи. Запишемо загальний вигляд блочної структури даної матриці:

$$M = \begin{bmatrix} EO \\ RW \end{bmatrix}, \tag{9}$$

де Е – одинична підматриця, порядок якої збігається з числом поглинаючих станів;

W – квадратна підматриця ймовірностей переходів на множині неповоротних станів;

R – прямокутна підматриця переходів з неповоротних станів в поглинаючі;

O – нульова підматриця. Якщо число загальних станів – 17, з них 2 – поглинаючі стани, тоді підматриці мають наступний порядок:

$$M_{17*17},\ E_{15*15},\ W_{2*2},\ R_{2*15},\ O_{15*2}.$$

Результати обчислень умовних імовірностей отримання очка при подачі наведені у таблиці 1:

Таблиця 1 – Таблиця умовних імовірностей отримання очка при подачі

		гравець А				
		0	15	30	40	Гейм
гравець В	0	0.962890674	0.896661648	0.737136081	0.419136599	0
	15	0.98267337	0.944312142	0.832122939	0.544333246	0
	30	0.994131919	0.977823203	0.918086095	0.706926293	0
	40	0.999003354	0.995666754	0.981159802	0.918086095	
	Гейм	1	1	1		

Ці данні показують, що гравець В зможе виграти гейм при подачі з ймовірністю 98%. Що і сталося в реальному матчі.

Знайдемо умовні ймовірності отримання очок кожним із гравців та ймовірність виграшу гейма гравцем В за умови, що він приймає подачу. Для цього побудуємо таблицю умовних ймовірностей та візуалізуємо її марківською моделлю гейму.

Результати обчислень умовних імовірностей отримання очка при прийомі наведені у таблиці 2:

Таблиця 2 – Таблиця умовних імовірностей отримання очка при прийомі подачі

		гравець А				
		0	15	30	40	Гейм
гравець В	0	0.285501053	0.173073318	0.081225575	0.022444053	0
	15	0.447287306	0.30524446	0.165813619	0.054741592	0
	30	0.651690425	0.505888841	0.325648973	0.133516079	0
	40	0.86150246	0.765258408	0.602132894	0.325648973	
	Гейм	1	1	1		

Ці данні показують, що гравець В зможе виграти гейм при прийомі подачі з ймовірністю 28 %. Тобто з великою ймовірністю, що складає 72 %, він програє цей гейм. Що і сталося в реальному матчі.

Візуалізацію ланцюга за умови, що гравець В приймає подачу наведено на рис.2:

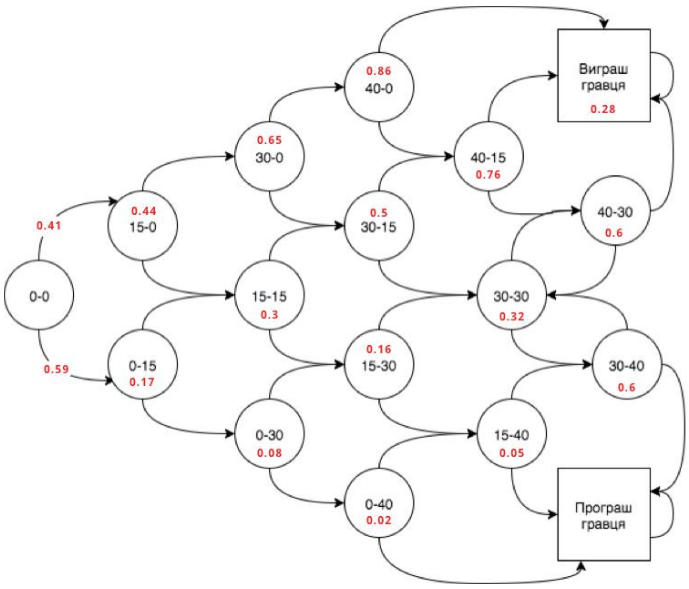

Рисунок 2 – Марківська модель гейму в матчі за умови, що гравець В приймає подачу

Візуалізацію ланцюга за умови, що гравець В подає наведено на рис.3:

Порівнявши отримані данні з даними статистики матчу можемо вважати, що даний спосіб аналізування тенісного гейму може бути використаний для практичного застосування та давати результати, що цілком узгоджуються з реальними даними. При цьому похибка є невеликою.

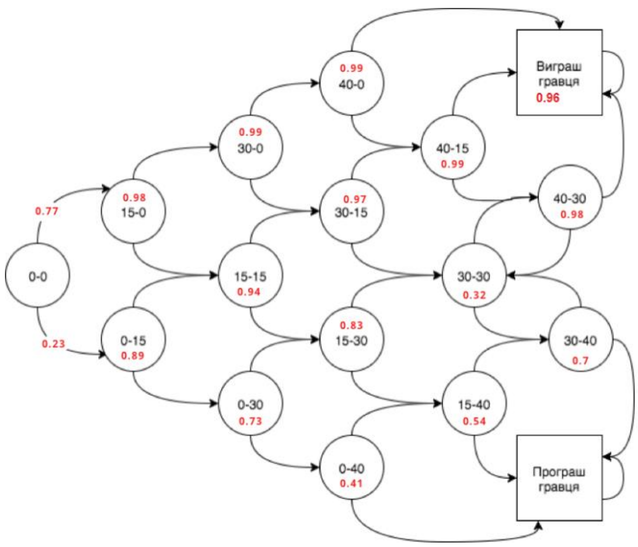

Рисунок 3 – Марківська модель гейму в матчі за умови, що гравець В подає

Висновки

У даній роботі було розглянуто використання ланцюга Маркова до задачі прогнозування тенісних турнірів. Було показано основні принципи побудови та умови, за яких побудова може бути виконана. Для покращення роботи даного методу можна застосовувати модифікації, пов'язані зі способом визначення ймовірності виграшу очка гравцем при подачі та на прийомі.

Використання марківських ланцюгів для прогнозування результатів тенісних турнірів не можна вважати досконалим методом, оскільки при цьому абсолютно ігнорується безліч факторів, що впливають на гру: самопочуття гравця, наявність нещодавніх травм, графік тренування, тип покриття корту тощо. Але урахування різноманітних статистичних даних спортсменів дає змогу ще більше наблизити до реального стану речей дану ідеалістичну модель і побудувати ланцюги Маркова, які точніше відображають сутність тенісного матчу.

Література

1. F.J.G.M. Klaassen and J.R. Magnus. Are Points in Tennis Independent and Identically Distributed? Evidence From a Dynamic Binary Panel Data Model. Journal of the American Statistical Association, 96:500–509, 2001.
2. M. Sipko. Machine Learning for the Prediction of Professional Tennis Matches. Technical report, Imperial College London, London, 2015.
3. T. Barnett and S. R. Clarke. Combining player statistics to predict outcomes of tennis matches. IMA Journal of Management Mathematics, 16:113–120, 2005.
4. J.A. O'Malley. Probability Formulas and Statistical Analysis in Tennis. Journal of Quantitative Analysis in Sports, 4(2), 2008.
5. W. J. Knottenbelt, D. Spanias, and A. M. Madurska. A common-opponent stochastic model for predicting the outcome of professional tennis matches. Computers and Mathematics with Applications, 64:3820–3827, 2012.
6. A. M. Madurska. A Set-By-Set Analysis Method for Predicting the Outcome of Professional Singles Tennis Matches. Technical report, Imperial College London, London, 2012.
7. ATP World Tour: Tennis Statistics. [Електронний ресурс] – Режим доступу: http://www.atpworldtour.com/en/players/
8. Australian Open (м) 2018. [Електронний ресурс] – Режим доступу: https://www.championat.com/tennis/_grandslam/13109/match/669971.html#text-all

Рецензенти:
Гамеляк І.П., д-р техн. наук, НТУ (Київ)
Скобелєв В.Г., д-р техн. наук, д-р фіз.-мат. наук, професор, провідний науковий співробітник Інституту кібернетики ім. В.М.Глушкова, Київ, Україна.

Reviewers:
Hameliak I.P., Dr.Tech.Sci., NTU (Kyiv)
Skobeliev V.H, Dr.Tech.Sci., Dr.Phys.Math.Sci., professor, senior researcher, Hlushkov Institute of Cybernetics (Kyiv)

Стаття надійшла до редакції: **08.09.2017 р.**