

УДК 004.89

КАСКАДИРОВАНИЕ ИСКУССТВЕННЫХ НЕЙРОННЫХ СЕТЕЙ ДЛЯ ПРОГНОЗИРОВАНИЯ РЕЗУЛЬТАТОВ ВСЕМИРНОЙ УНИВЕРСИАДЫ 2021 (2022)

Крутиков А.К., Страбыкин Д.А., Подковырин В.Д.
*ФГБОУ ВО «Вятский государственный университет», Киров,
e-mail: yadrodisk@yandex.ru, strabykin@vyatsu.ru, usr00321@vyatsu.ru*

В статье описывается эксперимент по прогнозированию общего количества медалей, количества золотых и серебряных медалей, а также места в неофициальном командном зачете XXXI Всемирной летней Универсиады 2021 (2022) для команды Российской Федерации. Описывается построение фрагментированной обучающей выборки. Описывается построение каскада нейросетевых модулей для осуществления прогнозирования, приводится пример построения каскада на основе разработанной фрагментированной структуры обучающей выборки. Эксперименты производятся с использованием как отдельных моделей нейронных сетей, так и с построением каскада нейросетевых модулей в различных конфигурациях. Рассматриваются проблемы, выявленные в ходе проведения экспериментов и пути их решения, к которым относится поиск основных и наиболее важных параметров, влияющих на результат прогноза, автоматизация выделения фрагментов выборки, детализация и предобработка выборки. Производится анализ результатов прогнозирования, результаты приводятся в таблицах, фрагмент обучающей выборки приведен на графике. Целью проводимых исследований являются прогнозирование результатов НКЗ с использованием многоярусного каскада нейросетевых модулей, а также оптимизация и повышение точности процесса планирования спортивной подготовки студенческих сборных команд, на основе результатов прогноза. Рассматриваются перспективы дальнейшего использования подхода, основанного на каскадировании нейросетевых модулей, возникающие при этом проблемы и возможности.

Ключевые слова: искусственная нейронная сеть, прогноз, спортивный результат, нейрон, прогнозирование, алгоритм обучения, обучающая выборка, обобщенно-регрессионная нейронная сеть, нейронная сеть с радиально-базисными функциями, каскадная нейронная сеть, каскадная структура, Всемирная Универсиада, неофициальный командный зачет, каскад

CASCADING OF ARTIFICIAL NEURAL NETWORKS TO PREDICT THE RESULTS OF THE WORLD UNIVERSIADE 2021 (2022)

Krutikov A.K., Strabykin D.A., Podkovyrin V.D.
Vyatka State University, Kirov, e-mail: yadrodisk@yandex.ru, strabykin@vyatsu.ru, usr00321@vyatsu.ru

The article describes an experiment to predict the total number of medals, the number of gold and silver medals, as well as a place in the unofficial team standings of the XXXI World Summer Universiade 2021 (2022) for the team of the Russian Federation. The construction of a fragmented training sample is described. The construction of a cascade of neural network modules for forecasting is described, an example of constructing a cascade based on the developed fragmented structure of the training sample is given. Experiments are carried out using both separate models of neural networks and with the construction of a cascade of neural network modules in various configurations. The problems identified during the experiments and ways to solve them are considered, which include the search for the main and most important parameters that affect the result of the forecast, automation of selection of sample fragments, detailing and preprocessing of the sample. The results of forecasting are analyzed, the results are shown in tables, a fragment of the training sample is shown on the graph. The purpose of the research is to predict the results of the NCZ using a multi-tiered cascade of neural network modules, as well as to optimize and improve the accuracy of the planning process of sports training of student national teams, based on the results of the forecast. The prospects of further use of the approach based on cascading of neural network modules, the problems and opportunities arising in this case are considered.

Keywords: artificial neural network, prediction, sports result, neuron, prediction, learning algorithm, training sample, generalized regression neural network, neural network with radial-basis functions, cascade neural network, cascade structure, World University Games, unofficial team competition, cascade

Всемирная летняя Универсиада – международные и национальные спортивные соревнования среди студентов, проводимые Международной федерацией университетского спорта (FISU). Всемирная Универсиада проводится раз в два года. На предстоящих состязаниях XXXI Летней Универсиады 2022 (2021) будет представлено 18 видов спорта [1], разыграно 268 комплектов наград. Сборная РФ (ранее СССР) принимает участие в Универсиадах с начала их проведения в 1959 г.

Выступление национальных студенческих команд оценивается аналогично выступлениям национальных команд на международных мультиспортивных мероприятиях, основной мерой такой оценки является место в неофициальном командном зачете (НКЗ). Место оценивается по основной версии – согласно количеству золотых медалей. Согласно второй распространенной версии НКЗ оценивается по общему количеству медалей вне зависимости от их достоинства. Подготовкой националь-

ной сборной в течение двух лет занимается Российский спортивный студенческий союз (РССС «Буревестник») [2]. Результаты выступления оцениваются представителями РССС, студенческих спортивных лиг и объединений. Информация широко тиражируется в СМИ. Стоит отметить что предстоящие игры Всемирной Универсиады (ВУ) были перенесены на один год, с 2021 г. на 2022 г., по причине пандемии COVID-19. Цикл подготовки команд был продлен, привычные подготовительные мероприятия перенесены или отменены. Результат НКЗ ВУ будет оцениваться специалистами национальных федераций по видам спорта и представителями РССС.

Медальный план оценивается и ставится перед студенческой сборной представителями РССС. Осуществление планирования медального плана требует работы экспертов, анализа динамики выступлений на предварительных турнирах, анализа ряда факторов, влияющих на состояние команды и отдельных спортсменов. Анализ факторов предполагает постоянную обработку данных с использованием как мнений экспертных групп [3], так и инструментов ИТ-технологий, включая направления искусственного интеллекта (ИИ) [4].

Среди таких средств можно выделить экспертные системы на основе логического вывода [5], технологии обработки больших данных, аппарат математической статистики, а также искусственные нейронные сети (ИНС), являющиеся распространенным средством прогнозирования [6].

Авторы данной работы неоднократно производили прогноз аналогичных спортивных событий, в том числе Всемирной зимней Универсиады 2019 [7] с использованием отдельных моделей ИНС. Среди используемых моделей рассматривались сети с радиальными функциями, каскадные сети прямого распространения, линейные нейронные сети и нейронные сети векторного квантования сигналов. Эксперименты проводились в том числе с использованием различных вариаций двухъярусного каскада нейросетевых модулей [8]. Одной из проблем является невозможность учета в выборке набора параметров, результаты которых прогнозируются в тот же период времени и влияют на основной результат прогноза. При добавлении таких параметров в выборку структура выборки изменяется, а следовательно, изменяется структура некоторых моделей ИНС, появляется необходимость переобучения сетей.

Авторами данной работы для решения этой проблемы применяется многоярусный каскад нейросетевых модулей, где промежу-

точные результаты, оказывающие влияние на результат прогнозирования, генерируются динамически, между ярусами каскада.

Объектом прогнозирования в данной статье является выступление сборной команды РФ на XXXI Летней Универсиаде 2022.

Целью проводимых исследований являются прогнозирование результатов НКЗ с использованием многоярусного каскада нейросетевых модулей, а также оптимизация и повышение точности процесса планирования спортивной подготовки студенческих сборных команд, на основе результатов прогноза.

Материалы и методы исследования

В работе [8] описывается построение двухъярусного каскада для прогнозирования результатов крупного мультиспортивного события. На первом ярусе каждый нейросетевой модуль производит промежуточный результат прогноза, с использованием отдельной модели нейронной сети, при этом каждый из модулей принимает одну логически разделенную группу параметров (условно социальную или условно спортивную), выраженную десятичными числами. Проведенные с двухъярусным каскадом эксперименты показали приближенные значения к верным результатам, с использованием каскада однотипных сетей, по сравнению с одной ИНС той же модели.

Обучающая выборка фрагментируется в соответствии с логическими взаимосвязями данных, данные могут дублироваться или отличаться полностью, в зависимости от логического назначения фрагмента выборки.

На первом этапе произведены эксперименты с отдельными моделями нейронных сетей. В данном случае использованы нейронные сети с радиальными функциями – обобщенно-регрессионная нейронная сеть (GRNN) и радиально-базисная нейронная сеть (RBF).

Для обучения данных сетей формируется обучающая выборка, включающая в себя результаты выступления сборных команд РФ начиная с 1992 г. (подготовка к Универсиаде 1993). Выборка содержит социальные и спортивные параметры, которые влияют на итоговый результат прогноза. Для формирования выборки использован тот же принцип, что и в работе [8], а также [9]. Выборка содержит данные о количестве делегации сборной команды на играх, результатах выступления команд на предстоящих играх турнирах, фактор «родных стен», количестве жителей, ВВП на душу населения и т.д.

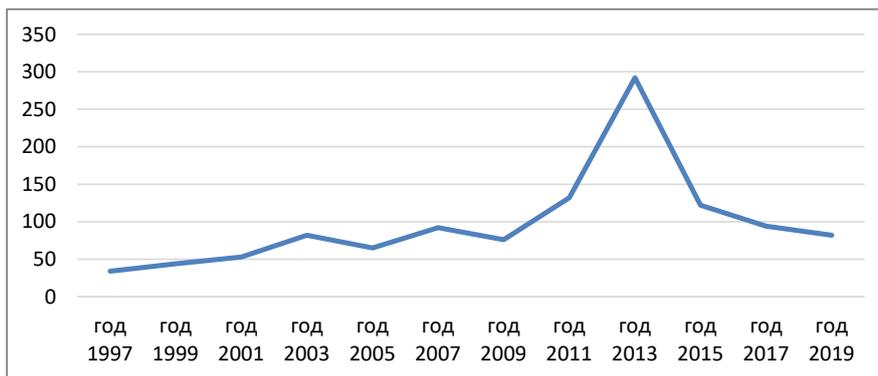


Рис. 1. Фрагмент результирующего вектора

Результирующий вектор содержит либо место в НКЗ по золотым медалям, либо общее место в НКЗ, либо отдельные результаты по общему количеству медалей, золоту или серебру. Бронзовые медали не учитываются, т.к. в НКЗ (по золоту) количество бронзовых медалей учитывается в последнюю очередь, при редком случае полного совпадения комбинации золото-серебро. Фрагмент результирующего вектора приведен на рис. 1.

В рабочем пространстве MATLAB с данными моделями нейронных сетей произведены эксперименты, результаты описываются в следующем разделе.

Часть результатов, полученных в ходе прогнозирования, могут оказывать влияние на основной прогнозируемый параметр (место сборной команды в общем зачете по общему количеству медалей или по количеству золотых медалей). В частности, речь идет о прогнозных значениях количества золотых, серебряных медалей и общего количества наград всех достоинств. При добавлении вновь полученных прогнозных значений в выборку необходимо перестроить структуру выборки, а с учетом выбранных моделей сетей при переобучении будет перестроена вся структура сети. В ряде программных и аппаратных реализаций, изменение структуры в ходе прогнозирования может быть недоступным и невозможным. В текущем эксперименте для отдельных моделей в обучающую выборку не включаются значения количества наград.

При применении каскада параметры, которые ранее не включались в выборку, формируются динамически в процессе прогнозирования, передаются между модулями каскада с одного яруса на другой. Данные значения являются промежуточными прогнозными результатами и одновременно прогнозоформирующими предикторами для модулей последующих ярусов. При об-

учении каскада каждый модуль обучается отдельно и включается в каскад после завершения процесса обучения.

Обучающая выборка фрагментируется следующим образом, на первом ярусе два блока параметров – социальные и спортивные, описанные ранее, в выборке, используемой для обучения отдельных моделей сетей. Во втором блоке параметров, для второго яруса, прогнозоформирующие значения количества золотых и серебряных медалей, одновременно являющиеся промежуточными прогнозными значениями для конкретного эксперимента. Третий ярус содержит третий блок прогнозоформирующих параметров, от которого зависит итоговый результат.

Структура обучающей выборки приведена на рис. 2.

Структура каскада нейросетевых модулей строится исходя из структуры фрагментированной обучающей выборки. Структура выборки приведена на рис. 3.

Авторами были спроектированы и реализованы в рабочем пространстве среды MATLAB несколько конфигураций каскада. Полученные программные модели занимают не более 8 Мбайт на жестком диске.

Результаты исследования и их обсуждение

Эксперименты проводились на модели системы спортивного прогнозирования, имеющей трехъярусную структуру, реализованную в среде MATLAB с использованием пакета Neural Network Toolbox. Нейронные сети RBF и GRNN были обучены с использованием рассмотренных в предыдущем разделе выборок. Результаты RBF-сети представлены относительно параметра сглаживания (SPREAD), результаты GRNN-сети представлены относительно значений целевой ошибки обучения (табл. 1).

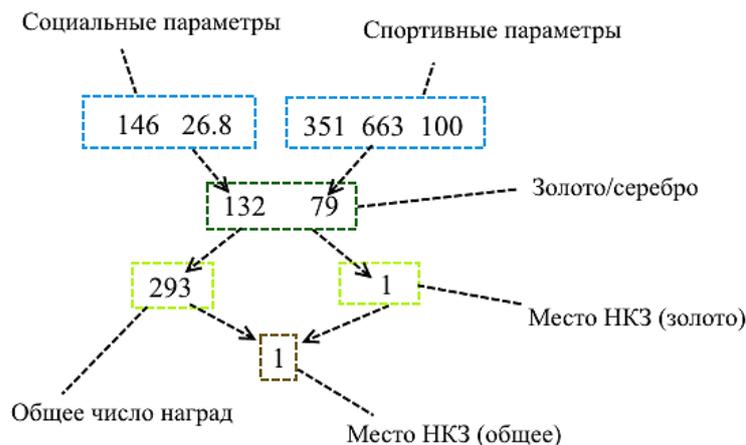


Рис. 2. Структура обучающей выборки

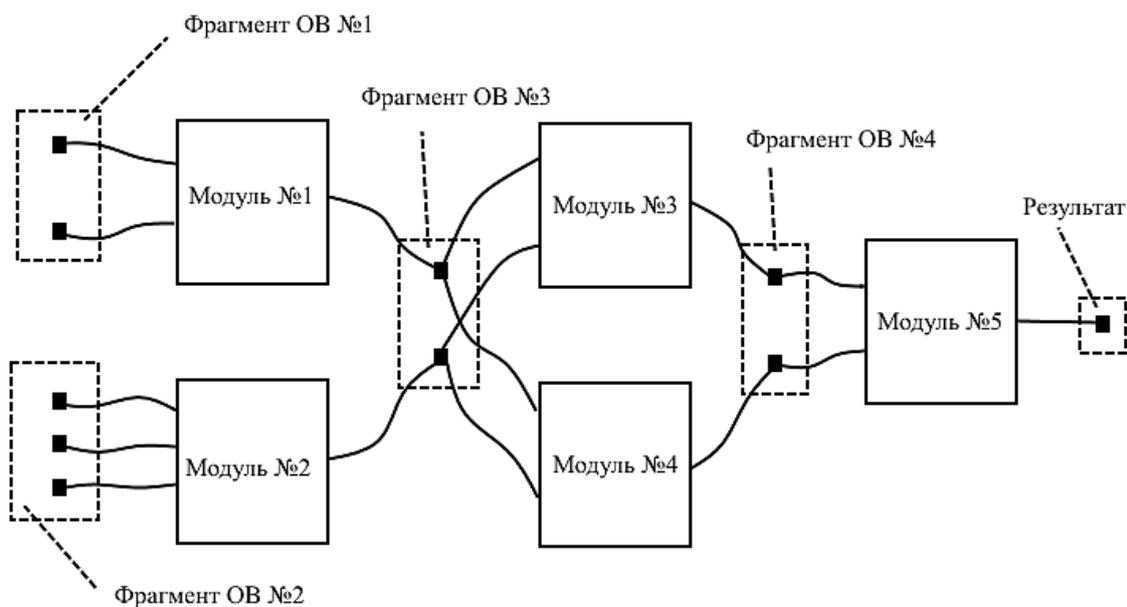


Рис. 3. Структура каскада нейросетевых блоков

Для нейронной сети с радиально-базисными функциями минимальная MSE для обучения получена при значении параметра SPREAD = 0.003.

При увеличении параметра более 3, MSE возрастает. Для обобщенно-регрессионной нейронной сети при увеличении параметра целевой ошибки более 10, MSE обучения начинает возрастать, результат прогноза при этом искажаются. Модуль на базе RBF показывает сомнительный результат – 1 место с 34 медалями, 10 из которых золотые, что маловероятно в условиях летней Универсиады.

Предлагается использование шести вариантов конфигурации каскада нейросетевых модулей. Отличие в предложенных конфигурациях заключается в расположении конкретных моделей ИНС в модулях определенных ярусов (рис. 4).

Параметры сглаживания SPREAD и целевой ошибки GOAL, соответственно в радиально-базисной сети [10] и обобщенно-регрессионной сети [11], используются равные 0.003 и 0.001 соответственно. Результаты эксперимента приведены в табл. 2.

В ходе эксперимента значение параметров не изменяется.

Таблица 1

Результат работы отдельных сетей

RBF-сеть			
SPREAD	Результат прогноза (место по золоту / общее место)	Результат прогноза (общее число медалей)	Результат прогноза (золото/серебро)
0.000003	1	34	10/14
0.00003	1	34	10/14
0.0003	1	34	10/14
0.003	1	34	10/14
0.03	1	34	10/14
0.3	1/1	34	10/14
GRNN-сеть			
Целевая ошибка	Результат прогноза (место по золоту / общее место)	Результат прогноза (общее число медалей)	Результат прогноза (золото/серебро)
0.001	1/2	122	35/39
0.01	1/2	122	35/39
0.1	1/2	122	35/39
1	1/2	122	35/39
2	1/2	122	35/39
50	1/1.23	117.26	30.26/31.11
100	0.92/1.42	114.92	29.37/29.73

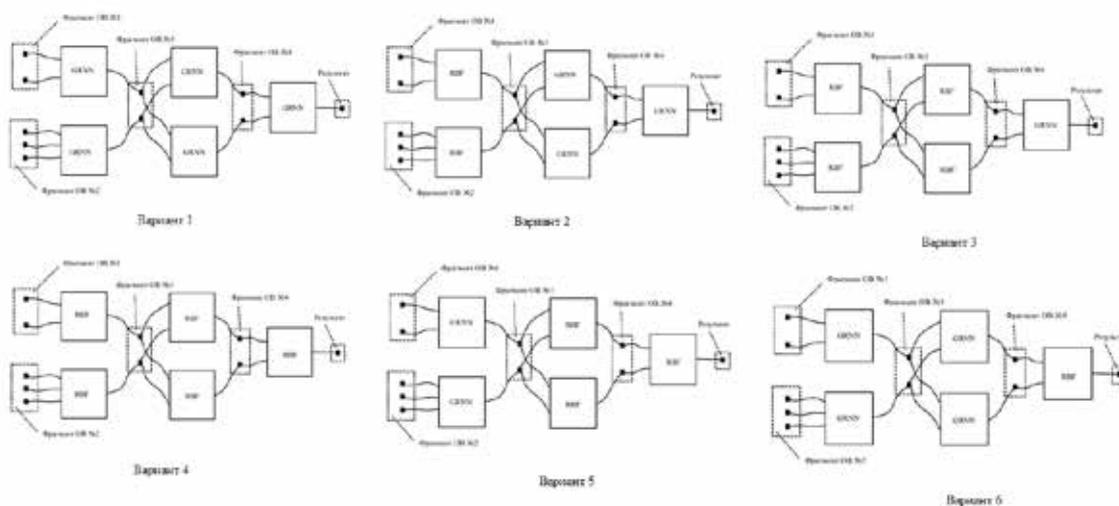


Рис. 4. Варианты конфигураций каскада

Вариант 5, очевидно, некорректно работает, следовательно, логический выбор распределения предикторов следует изменить. Варианты 2–4 демонстрируют маловероятные промежуточные данные. С точки зрения получения корректного прогноза стоит оценивать варианты 1 и 6. Точность прогноза возможно оценить после фактического завершения события.

Время обучения увеличивается в среднем на 1,55 с, поскольку модули в данном эксперименте обучаются одновременно. Фактическая работа каскада не превышает 6 с, что почти в 3,6 раза больше, чем работа отдельной обобщенно-регрессионной сети. В условиях данной задачи прогнозирования данное время не является существенным.

Таблица 2

Прогноз выступления сборной РФ

Каскад	Результат прогноза (место НКЗ общее)	Промежуточный результат прогноза (место НКЗ золото)	Промежуточный результат прогноза (золото/серебро)	Промежуточный результат прогноза (общее количество медалей)
Вариант 1	2	1	35/24	76
Вариант 2	2	4	10/1.34	34
Вариант 3	2	0.99	10/1.34	35
Вариант 4	0.99	0.99	10/1.34	35
Вариант 5	0.99	0.99	35/24	35
Вариант 6	0.99	1	35/24	76

Заключение

Использование информационных технологий, в частности технологий на основе нейронных сетей, при решении задач прогнозирования перспективно и широко применимо.

По сравнению с ошибками обобщения «простых» моделей нейронных сетей (например, GRNN-сетей) ошибка обобщения при использовании каскада уменьшается, следовательно, повышается точность прогнозирования. Полученная разница во времени в задачах с долгосрочным и сверхдолгосрочным периодом прогнозирования не имеет большого практического значения.

Конечно, и время обучения каскада из нескольких нейросетевых модулей будет больше времени обучения отдельной сети. Но обычно данный этап не включается в общее время функционирования системы. Кроме того, обучение модулей каскада можно осуществлять не только последовательно, но и параллельно. Переобучение всего каскада осуществляется только при изменении структуры обучающей выборки. Переобучение отдельных модулей потребует при добавлении в обучающую выборку новых наборов данных. При параллельном обучении время будет определяться подготовкой к работе самого «долгообучаемо» модуля.

Определение ключевых предикторов в представленном варианте специализированного модуля прогнозирования используется при оценке предельной ошибки обобщения (точности прогнозирования) самим пользователем. Он в ручном режиме может исключить некоторые параметры выборки, несущественно влияющие на изменение ошибки. В дальнейшем реализация алго-

ритма определения ключевых предикторов будет автоматизирована с учётом определения предельной ошибки самой системой.

Список литературы

1. Всемирная Универсиада 2022 [Электронный ресурс]. URL: <https://www.fisu.net/sport-events/summer-world-university-games-events/chengdu-2021-fisu-world-university-games> (дата обращения: 11.04.2022).
2. РССС Буревестник [Электронный ресурс]. URL: <https://studsport.ru/> (дата обращения: 11.04.2022).
3. Курамшин Ю.Ф. Проблемы прогнозирования высших спортивных достижений // Ученые записки университета имени П.Ф. Лесгафта. 2005. № 18. С. 40–58.
4. Negnevitsky M. Artificial intelligence, a guide to intelligent systems. Reading, MA, USA: Addison Wesley, 2005.
5. Долженкова М.Л., Страбыкин Д.А., Чистяков Г.А., Мельцов В.Ю. Вывод следствий в исчислении предикатов с построением схемы вывода // Современные наукоемкие технологии. 2018. № 3. С. 47–54.
6. Альжанов А.К., Кутпанова З.А., Гайнуллина И.И., Алимбекова Н.А. Решение задач и общий подход к прогнозированию с помощью нейронных сетей // International Scientific and Practical Conference World science. ROST. 2017. Т. 2. № 3. С. 26–31.
7. Крутиков А.К. Прогнозирование результатов Всемирной Зимней Универсиады 2019 с использованием искусственных нейронных сетей // Научное обозрение. Технические науки. 2019. № 2. С. 29–33.
8. Крутиков А.К., Мельцов В.Ю., Страбыкин Д.А., Подковырин В.Д. Ассемблирование искусственных нейронных сетей для прогнозирования результатов XXIV Зимних Олимпийских Игр 2022 // Современные наукоемкие технологии. 2021. № 12–1. С. 45–51.
9. Не счесть золота? [Электронный ресурс]. URL: <https://www.forbes.ru/forbes/issue/2004-08/20514-ne-schest-zolota> (дата обращения: 30.10.2021).
10. Введение в нейронные сети GRNN и PNN и в практику MATLAB [Электронный ресурс]. URL: <https://russianblogs.com/article/13761933/> (дата обращения: 23.09.2021).
11. Прошкина Е.Н., Балашова И.Ю. Анализ и прогнозирование успеваемости студентов на основе радиальной базисной нейронной сети // Технические науки: традиции и инновации: материалы III Международной научной конференции (Казань, март 2018 г.). Казань: Молодой ученый, 2018. С. 24–28.