

УДК 004.89

## ФОРМИРОВАНИЕ ФРАГМЕНТИРОВАННОЙ ОБУЧАЮЩЕЙ ВЫБОРКИ В СПЕЦИАЛИЗИРОВАННОЙ СИСТЕМЕ ПРОГНОЗИРОВАНИЯ СПОРТИВНЫХ СОБЫТИЙ

Крутиков А.К., Клюкин В.Ю., Подковырин В.Д., Кислицына М.И.

*ФГБОУ ВО «Вятский государственный университет», Киров,*

*e-mail: yadrodisk@yandex.ru, klyukin@vyatsu.ru, usr00321@vyatsu.ru, stud115697@vyatsu.ru*

В данной работе описывается подход к построению фрагментированной обучающей выборки. На основе фрагментированной обучающей выборки строится каскад нейросетевых модулей, который является инструментом прогнозирования спортивных результатов в индивидуальных и командных спортивных дисциплинах. При фрагментации обучающей выборки ключевую роль играют логические зависимости между ее параметрами. Само детальное разбиение на фрагменты построено на наличии прямой логической и/или транзитивной зависимости между набором (фрагментом) параметров и результирующим вектором, являющимся частью зависимого набора (фрагмента) параметров. Приводятся схематические изображения логических зависимостей между фрагментами выборки. Описывается пример фрагментирования обучающей выборки для прогнозирования спортивного события, а также приводится результат эксперимента с использованием полученной фрагментированной выборки. Рассматриваются вопросы автоматизированного определения логических зависимостей, формирования фрагментов выборок, нормализация и автоматизированный сбор данных обучающей выборки. Формирование фрагментированной обучающей выборки позволяет формировать нейросетевую каскад, который, в свою очередь, является «решающей» частью специализированной системы прогнозирования спортивных результатов. В свою очередь, алгоритмы формирования обучающей выборки и обучения отдельных нейросетевых блоков служат частью блока, управляющего «решающей» частью системы.

**Ключевые слова:** искусственная нейронная сеть, прогноз, спортивный результат, нейрон, прогнозирование, алгоритм обучения, обучающая выборка, каскадная структура, каскад, спортивное событие, фрагментированная обучающая выборка, фрагмент, система прогнозирования

## FORMATION OF A FRAGMENTED TRAINING SAMPLE IN AN SPECIALIZED SPORTS EVENT FORECASTING SYSTEM

Krutikov A.K., Klyukin V.L., Podkovyirin V.D., Kislitsyna M.I.

*Vyatka state university, Kirov, e-mail: yadrodisk@yandex.ru, klyukin@vyatsu.ru,*

*usr00321@vyatsu.ru, stud115697@vyatsu.ru*

This paper describes an approach to constructing a fragmented training sample. A cascade of neural network modules is built on the basis of a fragmented training sample, which is a tool for predicting sports results in individual and team sports disciplines. When fragmenting a training sample, logical dependencies between its parameters play a key role. The detailed fragmentation itself is based on the presence of a direct logical and/or transitive dependence between the set (fragments) of parameters and the resulting vector, which is part of the dependent set (fragment) of parameters. Schematic images of logical dependencies between the sample fragments are given. An example of fragmentation of a training sample for predicting a sports event is described, and the result of an experiment using the resulting fragmented sample is also given. The issue of automated determination of logical dependencies, formation of sample fragments, normalization and automated data collection of the training sample is considered. The formation of a fragmented training sample allows you to form a neural network cascade, which in turn is a «crucial» part of a specialized sports performance forecasting system. In turn, the algorithms for forming a training sample and training individual neural network blocks are part of the control unit of the «decisive» part of the system.

**Keywords:** artificial neural network, prediction, sports result, neuron, prediction, learning algorithm, training sample, cascade structure, cascade, sports event, fragmented training sample, fragment, prediction system

Прогнозирование результатов спортивных событий требует оценки и анализа набора данных, влияющих на итоговый результат. Результатом прогноза являются численный результат или формализованное утверждение о наступлении некоторого исхода спортивного события. Набор данных, влияющих на итоговый результат прогноза, может быть слабо структурирован, иметь сильно отличающиеся по типу и размеру значения.

При прогнозировании с использованием средств машинного обучения такой набор данных называют обучающим набором, или обучающей выборкой [1]. Среди наиболее

популярных средств ИИ, применяемых в качестве инструмента прогнозирования, можно выделить генетические алгоритмы, экспертные системы на основе логического вывода [2], математические средства моделирования [3], искусственные нейронные сети [4, 5, 6], в том числе комбинированные решения на основе нейронных сетей [7] – ансамбли нейронных сетей [8].

В данной работе в качестве основного инструмента прогнозирования рассматриваются нейронные сети.

Целью проводимых исследований являются оптимизация и повышение точности

процесса прогнозирования в индивидуальных и командных видах спорта.

### Материалы и методы исследования

В работах [9, 10] описываются эксперименты по прогнозированию различных спортивных результатов: командных, мультикомандных, индивидуальных – с применением каскада нейросетевых модулей.

При прогнозировании в экспериментах применялись три типа выборок. Обучающая выборка (dataset, выборка) – совокупность данных, обработанная для использования алгоритмами машинного обучения (Machine Learning). Под использованием подразумеваются обучение модели и дальнейшее ее применение для решения реальных задач на основе готового и структурированного набора данных, представленного в табличном виде. Благодаря обучающей выборке реализуются настройка и оптимизация параметров алгоритма, вследствие чего модель обучается. Контрольная (тестовая) выборка не зависит от обучающей выборки и в идеальном варианте должна содержать в себе новые данные. Благодаря этому этапу происходит проверка качества обученной модели. Валидационная выборка также не зависит от обучающей выборки, помогает выбирать наилучшую обученную модель.

Для обучения каскада, предлагаемого авторами, стандартная обучающая выборка прямоугольного вида (двумерный массив) (рис. 1) не может быть использована, поскольку структура каскада предполагает применение ограниченных фрагментов.

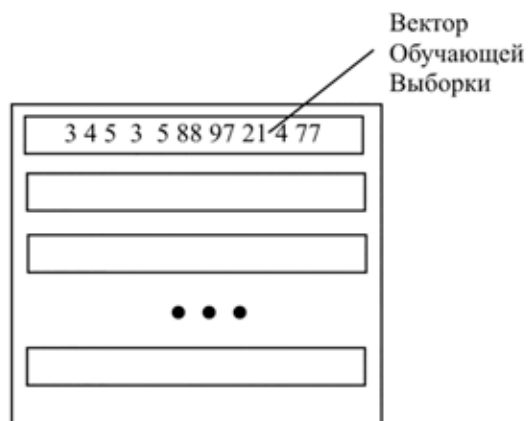


Рис. 1. Обучающая выборка (двумерный массив)

Каждая из строк двумерного массива содержит набор параметров одного «наблюдения» в один временной период. Набору таких параметров соответствует известный

элемент, известный в данный момент времени итоговый (профильный) результат.

При формировании фрагментированной структуры обучающей выборки двумерный массив данных разбивается на двумерные массивы меньшей размерности (рис. 2). Основным принципом разбиения является наличие логических зависимостей между элементами выборки, при условии, что элемент изначально не является итоговым результатом.

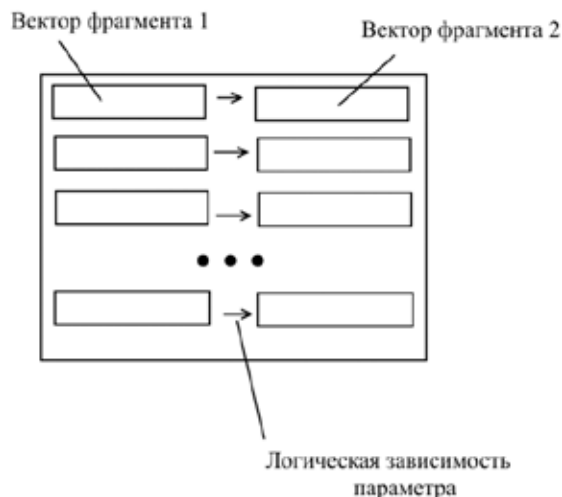


Рис. 2. Разбиение обучающей выборки на два фрагмента

Определение зависимости фрагментов обучающей выборки происходит на основе наличия логической зависимости между отдельным прямоугольным фрагментом и вектором и «столбцом» элементов из связываемого фрагмента. Количество связываемых фрагментов не ограничено. Логические связи могут быть выстроены исходя из предметной области, экспертного мнения или математической связи, выявленной с помощью отдельного блока анализатора.

Фрагменты выстраиваются в логически взаимосвязанную последовательность или последовательности (рис. 3).

Выделяются два типа параметров: предикторы – независимые переменные; целевые признаки – зависимые переменные, вычисляемые на основе предикторов.

Существует первичный набор исходных данных, называемый генеральной совокупностью. В процессе порождения данных из генеральной совокупности формируются выборки. Выборкой является конечное подмножество из элементов генеральной совокупности, проанализированное, которое, можно узнать и поведение генеральной совокупности.

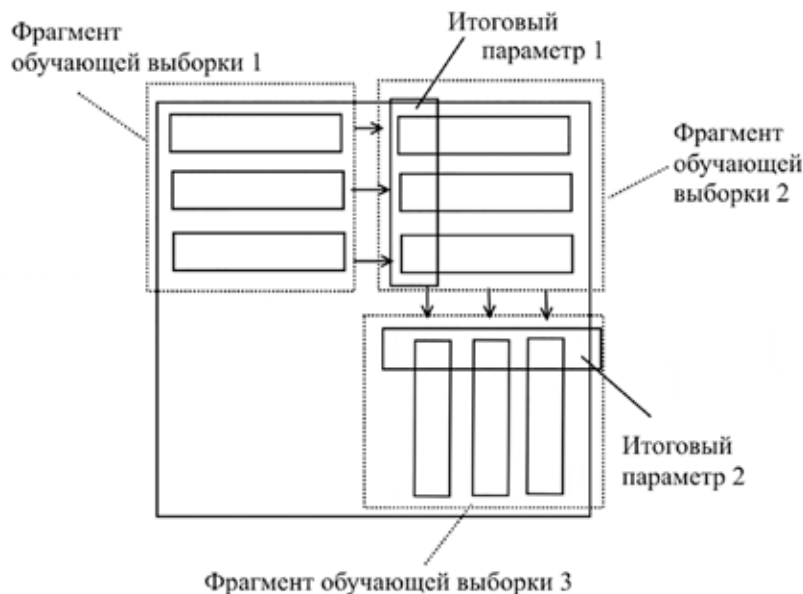


Рис. 3. Разбиение обучающей выборки на три фрагмента

Отдельные из параметров («столбцов») могут оказывать наибольшее влияние на результат или промежуточный результат (итоговый параметр в одном из фрагментов) прогноза. Такие параметры являются прогнозоформирующими предикторами, в работах [10-12] описан алгоритм определения прогнозоформирующих предикторов.

При обучении существующие фрагменты распределяются между выделенными уровнями нейросетевого каскада и являются обучающими выборками для отдельных нейросетевых модулей.

#### Результаты исследования и их обсуждение

На основе последовательностей определяются уровни нейросетевого каскада. Модули, содержащие сети, распределяются по ярусам каскада. Начиная со второго яруса, на каждый последующий ярус поступают данные промежуточных прогнозов, формирующие следующие прогнозные значения (параметры зависимых фрагментов). Количество прогнозоформирующих предикторов от яруса к ярусу может как уменьшаться, так и увеличиваться. В ряде задач при сокращении размеров обучающей выборки выделение прогнозоформирующих предикторов позволит уменьшить выборку без серьезных потерь в точности результата прогнозирования. Выявление зависимостей по ярусам каскада дает возможность определять те из предикторов, которые наибольшим образом воздействуют на текущие промежуточные результаты на выходе отдельных модулей.

Нейросетевые модули каскада обучаются отдельно, после чего соединяются друг с другом. В качестве эксперимента приведены результаты прогнозирования одного из международных спортивных событий – боксерского поединка. В эксперименте использовалась обучающаяся выборка, сформированная аналогично предыдущему примеру. Однако вектор выборки разбивается на две части. На основе результатов обработки исходных данных первым слоем получаем необходимые элементы выборки, поступающие на второй слой каскада. Нейронная сеть второго слоя обучалась отдельно. Пример разделения параметров во входной выборке и динамического формирования элементов выборок для последующих слоев приведен на рисунке 4.

Для реализации прогноза спроектирован каскад нейросетевых модулей. Каскад состоит из двух обобщенно-регрессионных нейронных сетей во входном (первом) слое и одной LVQ сети в заключительном модуле выходного (второго) слоя каскада [12]. Непосредственно результаты прогнозирования приведены в таблице.

Вектор (0,1) означает победу второго боксера, вектор (1,0) – победу первого боксера. В большинстве экспериментов система верно определила победителя. Определение логических взаимосвязей в обучающей выборке позволяет детализировать прогноз и динамически формировать промежуточные результаты, которые одновременно являются предикторами для последующих уровней каскада.

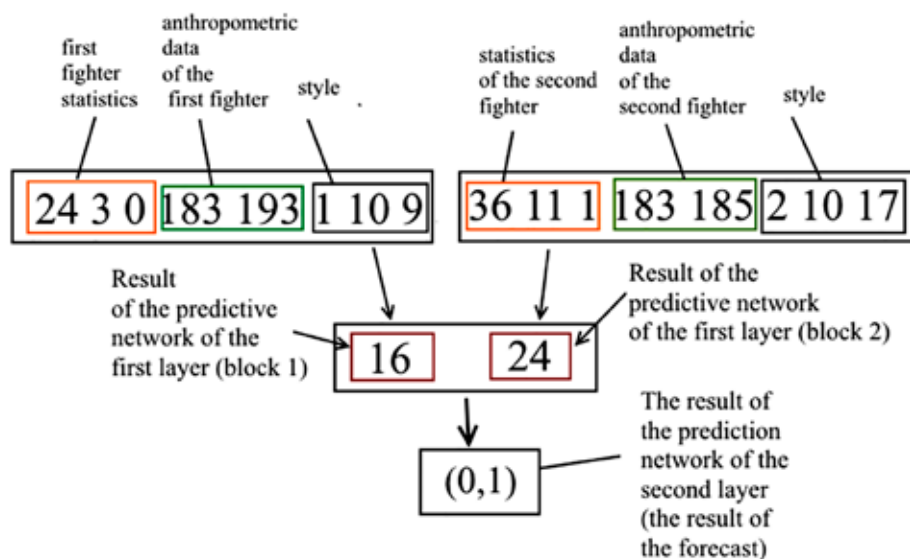


Рис. 4. Пример формирования обучающей выборки

#### Результат эксперимента

Количество нейронов второго яруса каскада (LVQ)	Результат прогноза
10	(1,0)
15	(1,0)
22	(0,1)
33	(0,1)
100	(0,1)
150	(0,1)
200	(0,1)
250	(0,1)

Нейросетевой каскада является «решающей» частью так называемой специализированной системы прогнозирования спортивных результатов. При этом фрагментирование выборки на основе логических зависимостей повышало точность прогнозирования как в мультиспортивных, так и в индивидуальных дисциплинах.

#### Заключение

Детализация прогноза путем динамического формирования вектора данных для последующих слоев каскада позволяет уменьшить объем информации, обрабатываемый последующими слоями, и увеличить точность прогноза.

Применение нейронных сетей для прогнозирования спортивных результатов в индивидуальных и командных дисциплинах перспективно и эффективно. Построение

нейросетевого каскада невозможно без выделения логических взаимосвязей в обучающей выборке. На данный момент тестируется отдельный модуль системы, автоматически определяющий не только прогнозоформирующие предикторы, но и наиболее сильные логические взаимосвязи между фрагментами выборки.

Продолжается тестирование различных вариантов нейросетевого каскада на основе новых вариантов фрагментированных обучающих выборок. Фрагменты обучающих выборок могут быть дополнены уникальными параметрами, которые не содержатся в исходном датасете.

Применение системы на основе нейросетевого каскада в качестве инструмента прогнозирования специалистами сферы физической культуры и спорта позволит повысить точность прогнозов, на основе которых можно выстраивать процесс планирования тренировки и подготовки спортсменов.

#### Список литературы

1. Обучающая выборка [Электронный ресурс]. URL: <http://www.machinelearning.ru/wiki/index.php?title=Выборка> (дата обращения: 12.04.2023).
2. Dolzhenkova M.L., Meltsov V.Y., Strabykin D.A. Method of consequences inference from new facts in case of an incomplete knowledge base // Indian Journal of Science and Technology. 2016. Vol. 9. No. 39. P. 100-413.
3. Емельянов С.В. Информационные технологии и вычислительные системы: Высокопроизводительные вычислительные системы. Математическое моделирование. Методы обработки информации. М.: Ленанд, 2012. 100 с.
4. Ростовцев В.С. Искусственные нейронные сети: учебник. Киров: Изд-во ВятГУ, 2014. 208 с.

5. Кудряшов В.А., Лезин И.А. Прогнозирование медицинских расходов рекуррентной сетью Элмана // Международная молодежная научная конференция, посвящённая 110-летию со дня рождения академика С.П. Королёва, 75-летию КуАИ-СГАУ-СамГУ-Самарского университета и 60-летию со дня запуска первого искусственного спутника Земли. Самара: Самарский национальный исследовательский университет имени академика С.П. Королева, 2017. 547 с.
6. Введение в нейронные сети GRNN и PNN и в практику MATLAB [Электронный ресурс]. URL: <https://russianblogs.com/article/13761933/> (дата обращения: 23.09.2021).
7. Осовский С. Нейронные сети для обработки информации / Пер. с польского. Д. Рудинского. М.: Финансы и статистика, 2002. 344 с.
8. Брюхнова В.О., Цуканова Н.И. Ансамбли нейронных сетей при прогнозировании объемов продаж в торговой сети // Вестник Рязанского государственного радиотехнического университета. 2018. № 66-1. С. 90-99.
9. Крутиков А.К. Механизм формирования обучающей выборки с выделением прогнозоформирующих предикторов для каскада нейронных сетей // Информационные технологии: материалы 60-й Междунар. науч. студ. конф. 10–20 апреля 2022 г. Новосибирск: ИПЦ НГУ, 2022. 106 с.
10. Meltsov V., Krutikov A., Strabykin D. Improving the Efficiency of Forecasting Sports Events Using a Cascade of Neural Networks // Lecture Notes in Networks and Systems this link is disabled. 2022. No. 461. P. 675–684.
11. Krutikov A.K., Meltsov V.Y., Strabykin D.A. Evaluation the Efficiency of Forecasting Sports Events Using a Cascade of Artificial Neural Networks Based on FPGA // Proceedings of ElConRus-2022. St. Petersburg: ETU LETI, 2022. P. 355–360.
12. Krutikov A.K., Meltsov V.Yu., Podkovyrin V.D. Neural Network of Vector Quantization as a Tool for Predicting Sports Achievements in Individual and Team Sports. Proceedings of the Volga Region Conference on Economics, Humanities and Sports (FICEHS 2019). Atlantis Press. P. 789-792.
13. Ясницкий Л.Н., Черепанов Ф.М. О возможностях применения нейросетевых технологий в политологии // Нейрокомпьютеры: разработка, применение. 2010. № 8. С. 47-53.