

Технологии нейронных сетей для прогнозирования успеваемости обучения студентов в электронной информационно-образовательной среде вуза

Токтарова В.И.

Марийский государственный университет (ФГБОУ ВО МарГУ)
г. Йошкар-Ола, Российская Федерация
ORCID: <https://orcid.org/0000-0002-3590-3053>
e-mail: toktarova@yandex.ru

Попова О.Г.

Марийский государственный университет (ФГБОУ ВО МарГУ)
г. Йошкар-Ола, Российская Федерация
ORCID: <https://orcid.org/0000-0002-2666-1005>
e-mail: olesya_popova10@mail.ru

Шашков О.В.

Марийский государственный университет (ФГБОУ ВО МарГУ)
г. Йошкар-Ола, Российская Федерация
ORCID: <https://orcid.org/0000-0002-0945-7096>
e-mail: olegs@citymed12.ru

В статье актуализируется проблема моделирования процесса обучения в условиях электронной информационно-образовательной среды на основе применения технологии нейронных с целью прогнозирования успеваемости обучения студентов. Обоснована актуальность использования технологий нейронных сетей для повышения качества учебно-педагогического процесса. Описан процесс проектирования и разработки прогностической модели успеваемости обучения студентов в электронной информационно-образовательной среде вуза, реализованной в рамках экспериментального исследования по учебной дисциплине «Основы программирования».

Ключевые слова: нейронные сети, прогнозирование успеваемости обучения, образовательные результаты, электронная информационно-образовательная среда, студенты, вуз.

Для цитаты:

Токтарова В.И., Попова О.Г., Шашков О.В. Технологии нейронных сетей для прогнозирования успеваемости обучения студентов в электронной информационно-образовательной среде вуза // Цифровая гуманитаристика и технологии в образовании (ДНТЕ 2022): сб. статей III Всероссийской научно-практической конференции с международным участием. 17–18 ноября 2022 г. / Под ред. В.В. Рубцова, М.Г. Сороковой, Н.П. Радчиковой. М.: Издательство ФГБОУ ВО МГППУ, 2022. 388–398 с.

Введение

Развитие электронного обучения и дистанционных образовательных технологий является обязательным компонентом формирования цифрового пространства знаний в соответствии со Стратегией развития информационного общества в Российской Федерации [4]. В федеральном проекте «Кадры для цифровой экономики» национальной программы «Цифровая экономика Российской Федерации» [1] обозначены и отражены цели и инициативы в области цифровизации сферы образования.

Одной из актуальных проблем высших учебных заведений является учебная успеваемость студентов. Это связано с высоким темпом развития технологий и повышением требований к профессиональным специалистам. Использование цифровых средств и реализация электронной информационно-образовательной среды (ЭИОС) становятся эффективным инструментом принятия решений по улучшению качества подготовки студентов [7], в т. ч. посредством прогнозирования успешности их обучения.

Достижения в области интеллектуального анализа данных и машинного обучения позволяют использовать информацию, накапливаемую системами управления обучением, для анализа зависимости успеваемости обучающихся от различных факторов. Для обработки больших объемов информации активно применяются нейронные сети. Под *нейронными сетями* подразумеваются вычислительные структуры, моделирующие биологические процессы, которые ассоциируются происходящими в человеческом мозге процессами. Они представляют собой распараллеленные системы, способные к обучению путем анализа положительных и отрицательных воздействий [2].

Использование моделей на основе нейронных сетей для прогнозирования успеваемости встречается в работах как отечественных, так и зарубежных исследователей. К примеру, в работе [3] нейронные сети применялись для прогнозирования группы риска по данным успеваемости студентов. Прогноз строился для студентов первого курса, в исследовании учитывались баллы ЕГЭ; иностранный язык, изучаемый в школе; наличие или отсутствие социальной стипендии; форма обучения; преподаватели по определенным дисциплинам и др. Авторы объясняют такой набор признаков тем, что они максимально позволяют описать образ первокурсника. В работе решалась задача бинарной классификации, модель определяла принадлежность студента к группе риска.

В частности, в исследовании [6] авторы использовали нейронные сети для прогнозирования итоговых оценок по курсу. Данные были

взяты из журналов логов, содержащих информацию об образовательных результатах за каждую неделю. Точность модели составила около 90 %. Результаты, полученные с помощью нейронных сетей, авторы сравнили с результатами регрессионного анализа и пришли к выводу, что нейронные сети эффективнее справились с поставленной задачей.

В работе [5] исследователи применили нейронной сети для прогнозирования успеваемости студентов не по отдельной дисциплине, а в целом, используя в качестве выходной переменной рейтинг студента в итоговом семестре. Студенты поделены на 2 класса: успешные с рейтингом выше 350 баллов и с низким рейтингом до 350 баллов. Обученная модель предсказывала к какому классу будет относиться абитуриент. В качестве входных данных учитывались такие показатели, как место жительства; школа, которую он окончил; выпускные оценки по физике и математике по окончанию школы; профессия отца и матери; сведения о семейном доходе. Авторы отметили, что разработанная модель с точностью 90 % предсказала успеваемость будущего студента.

Так, анализ и прогноз успешности обучения студентов рассматривались в исследованиях Е.В. Котовой, А.В. Колесниченко, Е.В. Прониной, Л.А. Колмогоровой, Е.В. Сеницына, Т.Ю. Быстровой, В.А. Ларионовой, Н.Ю. Фаткуллина и др. Исследования в области использования методов машинного обучения и нейронных сетей представлены в трудах А.В. Полбина, А.И. Павловой, Т.С. Станкевич, А.В. Баранова, В.Н. Клячкина, А.С. Кутузовой, И.А. Астраханцевой и др. Прогнозирование процессов в сфере образования на основе нейронных сетей приводятся в работах С.В. Русакова, О.Л. Русаковой, И.Ф. Ясинского, О.И. Федяева, Ф. Окубо, Т. Ямасита, А. Шимада, Х. Огата, А. Кехинде, Х. Гупта и др.

Целью данного исследования является проектирование, разработка и реализация модели прогнозирования успешности обучения студентов в условиях электронной информационно-образовательной среды вуза на основе технологии нейронных сетей.

Методы

Методологическую базу исследования составляет комплекс методов: теоретические (анализ нормативных источников, психолого-педагогической и специальной литературы; системный, структурно-функциональный и сравнительно-сопоставительный анализ; контент-анализ, систематизация, прогнозирование); эмпирические (наблюдение, опрос, тестирование, экспертная оценка.); математические (корреляционный анализ, методы математической статистики, математическое моделирование, технологии нейронных сетей).

Результаты

Информация, накапливаемая в процессе обучения, представляет собой массивы данных с большим количеством различных показателей. Их эффективно можно использовать для составления прогнозов об успеваемости обучающихся с помощью технологий нейронных сетей, преимуществом которых является возможность находить скрытые закономерности в больших данных, которые исследователю сложно обнаружить. Кроме того, нейронные сети:

- результативны при решении плохо формализованных или неформализованных задач;
- устойчивы к изменениям среды и характеру воздействия факторов;
- используются при неполноте или «зашумленности» используемых данных;
- применимы для решения задач, для которых характерны интуитивные решения.

Исследование прогнозирования успеваемости обучения студентов в электронной среде на основе нейронных сетей проводилось в несколько этапов.

На первом этапе были получены и предварительно обработаны данные о поведении/активности и промежуточных результатах обучения студентов в рамках дисциплины «Основы программирования», представленной в электронной информационно-образовательной среде вуза на базе в LMS Moodle. Выгрузка из курса содержала данные о времени, проведенном студентом в системе, количество кликов по каждому элементу курса, оценки за каждое задание по материалам лекций, оценки за тестовые задания и лабораторные работы. Дисциплина «Основы программирования» для студентов института цифровых технологий включала в себя 8 разделов, каждый из которых содержал конспект лекции и задание по ее содержанию, 1–2 лабораторные работы и 1–2 проверочных теста.

На втором этапе образовательные данные были проанализированы с целью выявления закономерностей. Корреляционный анализ показал, что итоговая оценка студента на курсе зависит не только от оценок, полученных им за задания, но также коррелирует с его онлайн-поведением на курсе. На рис. 1 приведена матрица корреляции, отражающая зависимость итоговой оценки от действий студента в ЭИОС (кликов по элементам курса).

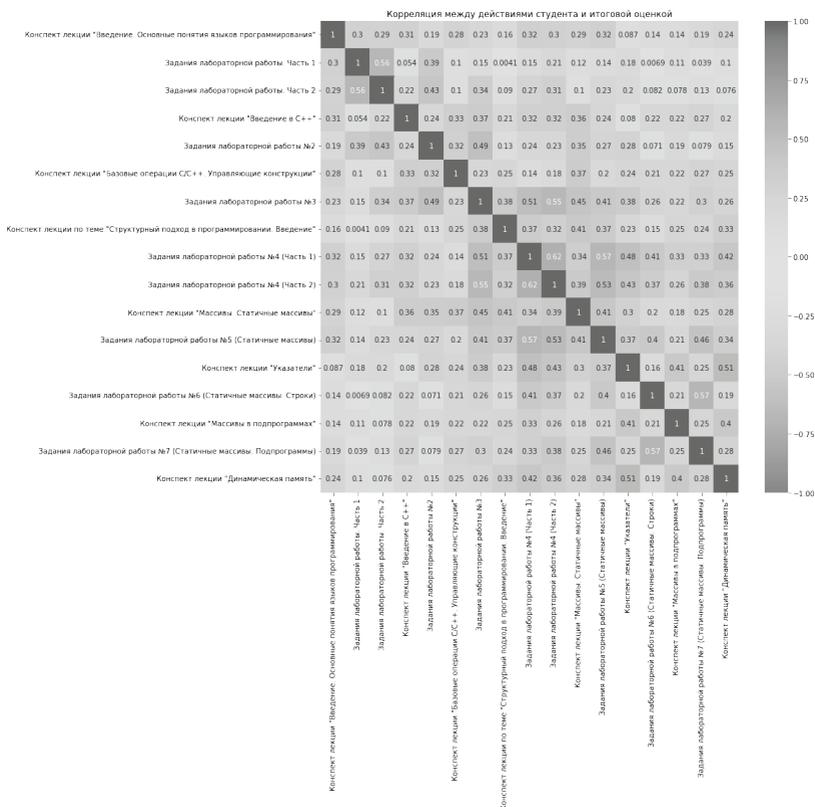


Рис. 1. Матрица корреляции между итоговой оценкой студента и его действиями на курсе

На третьем этапе после изучения структуры данных, результаты корреляционного анализа было принято решение использовать в модели прогноза в качестве переменной y итоговую оценку студента на курсе, а в качестве переменных x взять следующие показатели:

- x_1 — общее количество времени, проведенное в системе;
- x_2 — клики по конспекту лекции по теме «Введение. Основные понятия языков программирования»;
- x_3 — оценка за задание по конспекту лекции по теме «Введение. Основные понятия языков программирования»;
- x_4 — оценка за проверочный тест по теме «Введение. Основные понятия языков программирования»;
- x_5 — клики по заданиям лабораторной работы № 1 (часть 1);

- x_6 – клики по заданиям лабораторной работы № 1 (часть 2);
- x_7 – клики по конспекту лекции по теме «Введение в C++»;
- x_8 – оценка за задание по конспекту лекции по теме «Введение в C++»;
- x_9 – оценка за проверочный тест по теме «Введение в C++»;
- x_{10} – клики по заданиям лабораторной работы № 2;
- x_{11} – клики по конспекту лекции «Базовые операции C/C++» и т.д.

Всего в исследовании использовалось 36 факторных признаков. В данном случае, клики представляют собой переход к соответствующим элементам курса (конспекту лекции, лабораторной работе и др.). Количество кликов показывает, сколько раз студент открывал тот или иной элемент исследуемого электронного курса.

На четвертом этапе для прогнозирования было принято решение использовать метод многоклассовой классификации. Для построения модели итоговые оценки студентов были округлены до целых. Модель была реализована с использованием языка программирования Python при помощи библиотек: Keras и TensorFlow (для глубокого обучения), Pandas и NumPy (для обработки и анализа данных), Matplotlib и Seaborn (для визуализации данных) и Scikit-learn (для подготовки данных к построению модели).

Перед построением модели нейронной сети данные были стандартизированы путем вычитания среднего значения и деления на стандартное отклонение. Далее набор данных был разделен следующим образом: 80 % данных были использованы при обучении модели, 20 % – при тестировании. Для создания прогнозирующей успешности студента использовалась свёрточная нейронная сеть. Особенность такой архитектуры в том, что слой свёртки создает ядро, которое свертывается с входом слоя в едином пространстве (или временном) измерении для получения тензора выходных данных.

Модель состоит из трёх свёрточных слоёв (Conv1D). После первого и третьего слоя в модели обнуляется часть весов (слой Dropout). Далее следует слой, который уменьшает выборку, оставляя только самые важные признаки (MaxPooling), которые затем преобразуются в одномерные тензоры (слой Flatten) и передаются на полносвязные слои (Dense).

На пятом этапе производилось обучение нейронной сети и проверка точности модели. Обучением нейронной сети называется подбор входных весов для каждого нейрона таким образом, чтобы на выходе получить результат, максимально соответствующий ожидаемому. Точность и потери построенной модели свёрточной нейронной сети представлены на рис. 2.

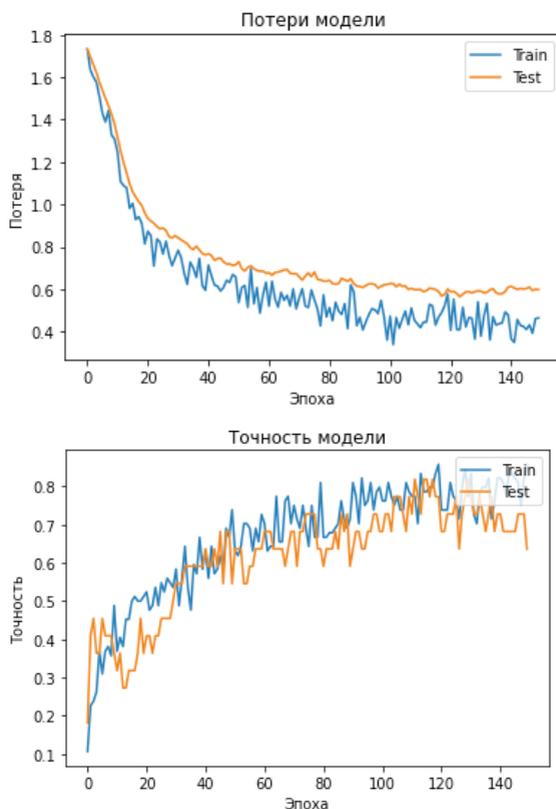


Рис. 2. Точность и потери на каждой эпохе обучения

Точность данной модели на тестовых данных составила 91,7 %. В рамках исследования был соотнесен прогноз на тестовых данных с реальными показателями. Исходя из этого, был сделан вывод, что данная модель даёт достаточно точный прогноз для использования предсказания успешности обучения студентов на курсе «Основы программирования» в условиях ЭИОС вуза.

На основе предсказанного значения можно вовремя выявить студента, который имеет высокую вероятность получения неудовлетворительной оценки.

Обсуждение

В ходе апробации и реализации модели было выявлено, что чем чаще студент обращается к конспекту лекции по конкретной теме,

тем лучше он выполняет задание по ней. Также у студентов, которые провели в системе большее количество времени за изучением учебного материала или его выполнением, итоговые оценки выше, чем у тех, у кого продолжительность присутствия на курсе была меньше. При этом график распределения оценок по заданиям выглядит следующим образом (рис. 3).



Рис. 3. Распределение оценок по заданиям

По данному графику можно выявить, какие задания вызывают у студентов наибольшую сложность и предпринять необходимые меры (представить подробное изложение материала темы, отнестись более тщательно к ее дидактическому оформлению, т.д.).

В результате реализации модели были получены следующие прогнозные значения для групп студентов: 5,9 % – «отлично», 23,7 % – «хорошо», 39 % – «удовлетворительно» и 31,4 % – «неудовлетворительно». Из полученного прогноза, с одной стороны, можно судить о сложности данного курса: около трети студентов имеют риск получить неудовлетворительную оценку. С другой стороны, можно сделать вывод о невыполнении более половины группы студентов проверочных заданий вовремя, оставляя на конец семестра, что также негативно может сказаться на их успеваемости.

Таким образом, данная модель обладает достаточно хорошей точностью прогнозирования и может использоваться на курсе «Основы программирования» и на других схожих по структуре учебных курсах. Предложенная модель позволяет предсказать итоговую оценку по дисциплине и выявить студентов с высоким риском получения неудовлетворительной итоговой оценки, тем самым способствуя устранению проблемы отчисления студентов за счет своевременного принятия необходимых мер.

Литература

1. Паспорт федерального проекта «Кадры для цифровой экономики» [Электронный ресурс] // Цифровая экономика 2024. URL: <https://digital.ac.gov.ru/poleznaya-informaciya/material/Паспорт-федерального-проекта-Кадры-для-цифровой-экономики.pdf> (дата обращения: 04.08.2022).
2. Позднеев Б.М., Кабак И.С., Суханова Н.В. Контроль знаний студентов на основе нейронных сетей // Открытое образование. 2011. № 6. С. 17–20.
3. Рузаков С.В., Рузакова О.Л., Посохина К.А. Нейросетевая модель прогнозирования группы риска по успеваемости студентов первого курса // Современные информационные технологии и ИТ-образование. 2018. № 4. С. 815–822.
4. Указ Президента РФ от 09.05.2017 г. № 203 «О Стратегии развития информационного общества в Российской Федерации на 2017–2030 годы» [Электронный ресурс] // Президент России. URL: <http://www.kremlin.ru/acts/bank/41919> (дата обращения: 08.04.2022).
5. Ясинский И.Ф., Семенова М.Б. Опыт прогнозирования успеваемости студентов при помощи нейросетевой технологии // Вестник ИГЭУ. 2007. № 4. С. 29–31.
6. Okubo F. et al. A neural network approach for students' performance prediction // The Seventh International Learning Analytics & Knowledge Conference. 2017. P. 598–599. doi: 10.1145/3027385.3029479
7. Toktarova V.I. Pedagogical management of learning activities of students in the electronic educational environment of the university: a differentiated approach // International Education Studies. 2015. Т. 8. № 5. С. 205–212. doi:10.5539/ies.v8n5p205

Информация об авторах

Токтарова Вера Ивановна, доктор педагогических наук, профессор кафедры прикладной математики и информатики, Марийский государственный университет (ФГБОУ ВО МарГУ), г. Йошкар-Ола, Российская Федерация, ORCID: <https://orcid.org/0000-0002-3590-3053>, e-mail: toktarova@yandex.ru

Попова Олеся Геннадьевна, магистрант 2 курса Института цифровых технологий, Марийский государственный университет (ФГБОУ ВО МарГУ), г. Йошкар-Ола, Российская Федерация, ORCID: <https://orcid.org/0000-0002-2666-1005>, e-mail: olesya_porova10@mail.ru

Шашков Олег Владимирович, директор центра искусственного интеллекта, Марийский государственный университет (ФГБОУ ВО МарГУ), г. Йошкар-Ола, Российская Федерация, ORCID: <https://orcid.org/0000-0002-0945-7096>, e-mail: olegs@citymed12.ru

Neural Network Technologies for Predicting Student Learning Achievement within eLearning Environment of the HEI

Vera I. Toktarova

Mari State University, Yoshkar-Ola, Russia
ORCID: <https://orcid.org/0000-0002-3590-3053>
e-mail: toktarova@yandex.ru

Olesya G. Popova

Mari State University, Yoshkar-Ola, Russia
ORCID: <https://orcid.org/0000-0002-2666-1005>
e-mail: olesya_popova10@mail.ru

Oleg V. Shashkov

Mari State University, Yoshkar-Ola, Russia
ORCID: <https://orcid.org/0000-0002-0945-7096>
e-mail: olegs@citymed12.ru

The article actualizes the problem of modeling the learning process within eLearning environment based on the using neural networks technology to predict student learning progress. The relevance of using neural network technologies to improve the quality of the educational and pedagogical process is substantiated. The process of designing and developing a predictive model of student learning progress within eLearning environment of the HEI, implemented by an experimental study on the academic discipline “Fundamentals of Programming”, is described.

Keywords: neural networks, predicting of learning achievement, educational results, eLearning environment, students, HEI.

For citation:

Toktarova V.I., Popova O.G., Shashkov O.V. Neural Network Technologies for Predicting Student Learning Achievement in the Electronic Information and Educational Environment of the University // *Digital Humanities and Technology in Education (DHTE 2022): Collection of Articles of the III All-Russian Scientific and Practical Conference with International Participation. November 17–18, 2022* / V.V. Rubtsov, M.G. Sorokova, N.P. Radchikova (Eds). Moscow: Publishing house MSUPE, 2022. 388–398 p. (In Russ., abstr. in Engl.).

References

1. Passport federal'nogo proekta «Kadry dlya tsifrovoy ekonomiki» [Elektronnyi resurs] [Passport of the federal project “Personnel for the Digital Economy”]. Tsifrovaya ekonomika 2024 [Digital Economy 2024]. URL: <https://digital.ac.gov.ru/poleznaya-informaciya/material/Pasport-federal'nogo-proekta-Kadry-dlya-tsifrovoy-ekonomiki.pdf> (Accessed: 04.08.2022).

2. Pozdneeve B.M., Kabak I.S., Sukhanova N.V. Kontrol' znaniy studentov na osnove nejronnykh setey [Control of students' knowledge based on neural networks]. *Otkrytoe obrazovanie = Open Education*, 2011, no. 6, pp. 17–20. (In Russ., Abstr. in Engl.).
3. Rusakov S.V., Rusakova O.L., Posokhina K.A. Neurosetevaya model' prognozirovaniya gruppy riska po uspevaemosti studentov pervogo kursa [A neural network model for predicting a risk group based on the progress of first-year students]. *Sovremennye informatsionnye tekhnologii i IT-obrazovanie = Modern information technologies and IT education*, 2018, no. 4, pp. 815–822. (In Russ., Abstr. in Engl.).
4. Ukaz Prezidenta RF ot 09.05.2017 g. № 203 «O Strategii razvitiya informatsionnogo obshchestva v Rossiiskoi Federatsii na 2017–2030 gody» [Elektronnyi resurs] [Decree of the President of the Russian Federation of May 9, 2017 No. 203 “On the Strategy for the Development of the Information Society in the Russian Federation for 2017–2030”]. Prezident Rossii [President of Russia]. URL: <http://www.kremlin.ru/acts/bank/41919> (Accessed: 08.04.2022).
5. Yasinskii I.F., Semenova M.B. Opyt prognozirovaniya uspevaemosti studentov pri pomoshchi neurosetevoi tekhnologii [The experience of students' progress forecasting using neuronet technology]. *Vestnik IGEU = Vestnik of Ivanovo State Power Engineering University*, 2007, no.4, pp. 29–31. (In Russ., Abstr. in Engl.).
6. Okubo F. et al. A neural network approach for students' performance prediction. *The Seventh International Learning Analytics & Knowledge Conference*, 2017, pp. 598–599. doi: 10.1145/3027385.3029479
7. Toktarova V.I. Pedagogical management of learning activities of students in the electronic educational environment of the university: a differentiated approach. *International Education Studies*, 2015, vol. 8, no. 5, pp. 205–212. doi:10.5539/ies.v8n5p205

Information about the authors

Vera I. Toktarova, Dr. of Pedagogical Sciences, Mari State University, Yoshkar-Ola, Russia, ORCID: <https://orcid.org/0000-0002-3590-3053>, e-mail: toktarova@yandex.ru

Olesya G. Popova, 2nd year master's student at the Institute of Digital Technologies, Mari State University, Yoshkar-Ola, Russia, ORCID: <https://orcid.org/0000-0002-2666-1005>, e-mail: olesya_popova10@mail.ru

Oleg V. Shashkov, director of the artificial intelligence center, Mari State University, Yoshkar-Ola, Russia, ORCID: <https://orcid.org/0000-0002-0945-7096>, e-mail: olegs@citymed12.ru