

АКТУАЛЬНЫЕ ВОПРОСЫ РАЗРАБОТКИ НЕЙРОСЕТЕВЫХ МОДУЛЕЙ ПРОГНОЗИРОВАНИЯ УСПЕВАЕМОСТИ ОБУЧАЮЩИХСЯ ВЫСШИХ УЧЕБНЫХ ЗАВЕДЕНИЙ

CURRENT ISSUES IN THE DEVELOPMENT OF NEURAL NETWORK MODULES FOR PREDICTING THE ACHIEVEMENT OF STUDENTS IN HIGHER EDUCATIONAL INSTITUTIONS

A. Belyakova
O. Romashkova
T. Ermakova
S. Chiskidov

Summary. The article is devoted to the study of the development of neural network modules that could be used in higher education institutions.

Keywords: educational organization, performance forecasting, data analysis, neural networks.

Белякова Анна Вячеславовна

Аспирант, ГАОУ ВО «Московский городской педагогический университет»
itwhitelight@mail.ru

Ромашкова Оксана Николаевна

Доктор технических наук, профессор, профессор, ФГБОУ ВО «Российская академия народного хозяйства и государственной службы при Президенте РФ» г. Москва
ox-rom@yandex.ru

Ермакова Татьяна Николаевна

Кандидат технических наук, доцент, ГАОУ ВО «Московский городской педагогический университет»
ermaktat@bk.ru

Чискидов Сергей Васильевич

Кандидат технических наук, доцент, профессор ФГБОУ ВО «Академия гражданской защиты МЧС России» г. Химки (Московская обл.)
chis69@mail.ru

Аннотация. Статья посвящена исследованию вопросов разработки нейросетевых модулей, которые могут использоваться в высших учебных заведениях.

Ключевые слова: образовательная организация, прогнозирование успеваемости, анализ данных, нейросети.

Введение

Задача прогнозирования успеваемости обучающихся сама по себе является широко востребованной на данный момент в учебном процессе как обучающимися, так и преподавателями, поскольку нейросетевой модуль может выявить незаметные пробелы в знаниях по определённым дисциплинам до того, как обучающийся неудовлетворительно сдаст экзамены или государственную итоговую аттестацию.

Однако при разработке и внедрении современных нейросетевых технологий прогнозирования успеваемости обучающихся высших учебных заведений может возникнуть ряд вопросов, на которые постарается дать краткий ответ данная статья. Для решения поставленных задач в интересах как руководства высших учебных заведений, так и разработчиков, внедряющих подобные технологии, следует проанализировать нюансы этапов разработки.

1. Универсальность нейросетевого модуля

Есть необходимость предварительно определиться с терминологией. Что же такое нейросетевые модули? По сути, это программное обеспечение, которое можно интегрировать в существующую информационную систему, чтобы в дальнейшем система могла решать дополнительные функциональные задачи, связанные с прогнозированием результатов. И уже на этапе интеграции возникает вопрос универсальности и гибкости подобного модуля. Иными словами, если в одной образовательной организации установлена система «Moodle» или система «1С.Университет», или даже в организации «1С.Университет» и «Moodle» связаны между собой, то, для интеграции нейросетевого модуля прогнозирования в него должен быть встроен интерфейс API (англ. Application Programming Interface — программный интерфейс приложения) для взаимодействия с указанными сервисами. Для этого можно использовать тот же самый язык программирования, который наиболее распространён при создании ней-

Таблица 1.

Результаты сравнительного анализа сервисов для вузов

Характеристика Система	Наличие web-версии	Ценовая политика	Конфиденциальность данных	Наличие собственного API	Создание отчётов
1С:Электронное обучение. Корпоративный университет	Нет	Тарифные планы	Возможность тонкой настройки прав доступа различных групп пользователей	Нет	Да
Moodle	Да	Тарифные планы, пробный период	Шифрование, настройка ролей	Да	Да
Canvas	Нет	Бесплатно	Шифрование, аудит безопасности	Да	Да
Teachbase	Да	Тарифные планы, пробный период	Шифрование, аудит безопасности	Да	Да
Google Класс	Да	Бесплатно	Шифрование	Да	Нет
Контур.Класс	Да	Тарифные планы, пробный период	Шифрование, аудит безопасности	Нет	Нет

росетей — Python (либо, при желании, Express — для Node.js). И если для создания нейросети в современных условиях достаточно разместить за символическую плату задачу на сайте kaggle.com, то для создания API можно воспользоваться фреймворком «Flask», который способен отправлять и принимать запросы для получения данных об успеваемости обучающихся. Таким образом возможно объединить нейросетевой модуль с основной системой, то есть с такими программами, как «Moodle», «1С.Университет» или др. Наиболее распространённые программы и сервисы (таблица 1) имеют локализацию на русском языке.

2. Составляющие современного нейросетевого модуля

Как уже было сказано, нейросетевой модуль должен включать в себя интерфейс API для взаимодействия с основной системой, но помимо этого, составляющие части программного обеспечения должны быть построены по определённым алгоритмам для анализа успеваемости обучающихся и загрузки дополнительных данных. Программное обеспечение должно иметь определённую архитектуру и отвечать стандартам защиты персональных данных [1].

Разработка программного обеспечения содержит определённые этапы. Самым первым этапом обычно является разработка функциональных и нефункциональных требований.

Функциональные требования могут быть сформулированы подобным образом:

- определение формата возвращаемых данных, и того, какие данные, помимо успеваемости, можно вывести для дальнейшего анализа. В основном данные возвращаются в формате «JSON», но если в данном учебном заведении используется какая-то специфическая система, то необходимо будет изучить документацию по конкретному сервису.

Кроме успеваемости, возможно получить данные об активности в сервисе или о преподавателях; — собственно прогнозирование результатов успеваемости обучающихся на основе выгруженных данных.

Нефункциональные требования здесь менее актуальны, но стандартно можно отметить следующие:

- защита персональных данных — обезличенная успеваемость. Поэтому для получения согласия на обработку персональных данных необходимо придерживаться высоких стандартов информационной безопасности;
- возможность обработки больших объемов данных при достаточно большой скорости, чтобы каждый пользователь мог оперативно получить свой прогноз;
- обеспечение работы в операционной системе «Windows» и в одной из unix-подобных систем.

Следующим этапом после определения требований относительно источника является решение вопроса о наличии БД (базы данных) и определение объема обрабатываемых данных, обеспечивающих корректную работу нейросетевого модуля прогнозирования. Необходимо также разработка наилучшей архитектуры БД и форматов выходных данных.

Выгруженных из основной системы данных может хватить, однако при необходимости нейросетевой модуль прогнозирования успеваемости обучающихся может обладать собственной БД. Это может потребоваться при, к примеру, уже упомянутой связке «1С.Университет» и «Moodle». Наилучшим вариантом будет загрузка необходимых данных для анализа в преобразованном виде в БД. В качестве БД может использоваться стандартный вариант, — подойдет как «PostgreSQL» или «MySQL», так и «SQLite». Менее предпочтительна документно-ориентированная база, подобная «MongoDB». Поскольку

последняя известна, но не работает в России, можно использовать, к примеру, «Apache Cassandra», а для обеспечения нефункционального требования в области быстрой работы БД для ускорения можно дополнительно использовать «Redis» как кэширующую базу, которая функционирует в ОЗУ, — сама же она базируется в операционной памяти нереляционной БД, чьи данные хранятся в более традиционных для программирования структурах данных, объектах, построенных по типу ключ-значение.

Необходимость создания БД, скорее всего, возникнет, поскольку в основном данные поступают в систему в формате JSON, и в таком формате их тоже можно использовать для анализа нейросетью, но многие БД функционируют гораздо эффективнее, и работа с данными будет проще. Если всё же выбрана реляционная БД, что предпочтительнее, то на следующем этапе необходимо обеспечить хранение данных об успеваемости в удобной для обработки нейросетью форме. При наличии дополнительных данных, которые удалось выгрузить из сервиса, необходимо сохранять и их, как потенциальные косвенные факторы, влияющие на успеваемость [2]. Чтобы нейросетевой модуль прогнозирования давал рабочие прогнозы, база данных должна быть не только хорошо спроектирована, но и готова к обработке больших объемов данных, как, собственно, и сама нейросеть.

Помимо выделения достаточного объема серверов в зависимости от того, насколько большой объем данных об успеваемости выбран для прогнозирования (ри-

сунк 1), необходимо установить соответствующую БД на серверы.

Для создания самой нейросети необходимо разработать её модель, то есть выбрать наиболее подходящие для прогнозирования успеваемости алгоритмы и архитектуру (рисунок 2). Помимо этого, необходимо произвести обработку и предобработку данных об успеваемости, какого бы вида и объема они ни были, что является в целом стандартной задачей по удалению ошибок, приведению типов к одному стандарту, удалению дубликатов и так далее [3].

Поскольку никаких особенностей решения данной задачи нет, то более подробное описание не имеет смысла, если, как и рекомендуется, был выбран язык Python, — достаточно даже библиотеки типа «Pandas» для подобной обработки, без дополнительного использования нейросетей. Единственное, что действительно стоит отметить: данный этап обязателен, если в существующей системе было достаточно ошибок и пропусков в данных, то прогноз нейросети может оказаться неверным.

После получения и обработки данных, выбора подходящей модели обучения нейросети необходимо для обеспечения её производительности и корректной работы проверить её работу при использовании всё того же языка Python в библиотеке «TensorFlow». Можно реализовать LSTM рекуррентную нейронную сеть и произвести настройку слоёв, выбрать функцию активации,

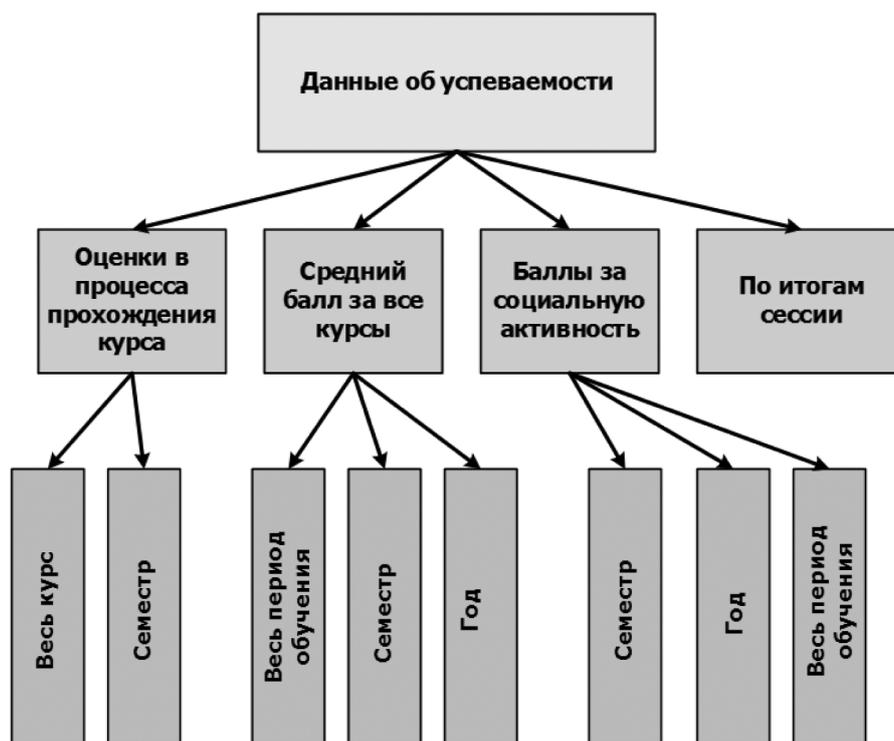


Рис. 1. Типы и объемы данных об успеваемости обучающихся

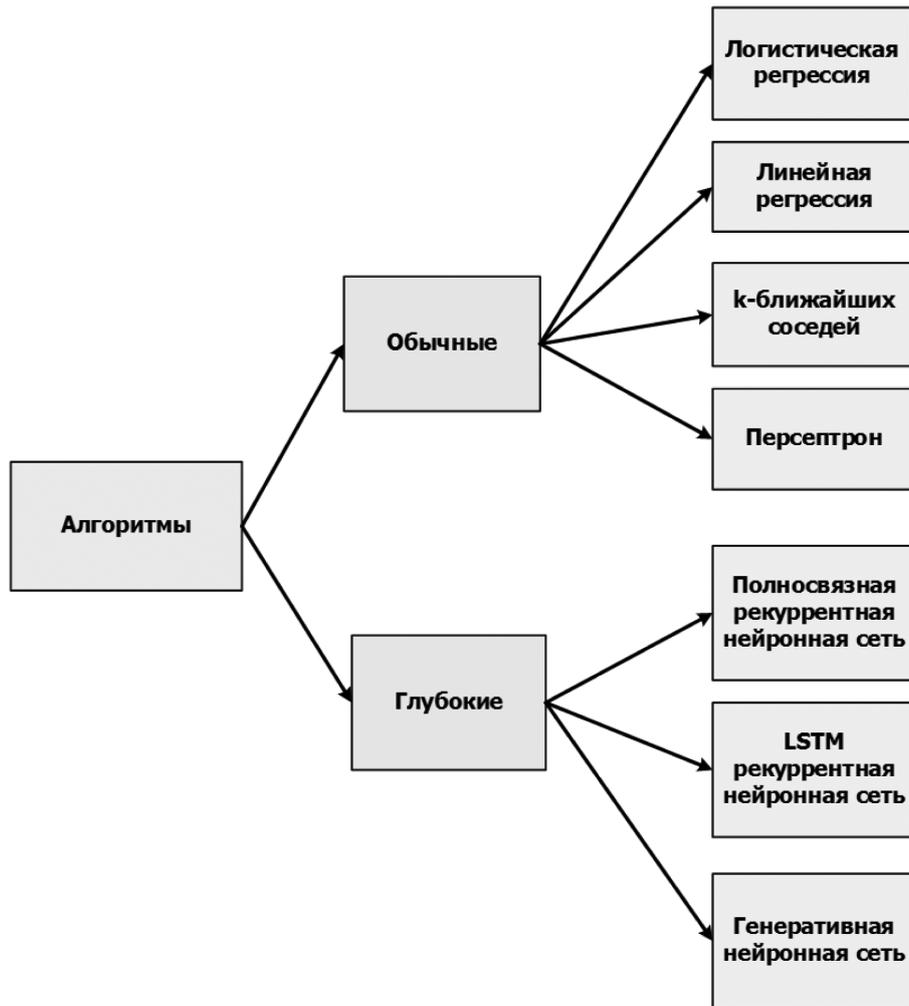


Рис. 2. Алгоритмы функционирования нейросетевого модуля

количество шагов, признаков, размер выходного слоя и произвести прочие настройки, для наиболее эффективной работы нейросети в нейросетевом модуле. Внутри самой LSTM также существуют функции активации на разных этапах работы, где гиперболический тангенс сжимает значения, а сигмоида преобразует их в вероятности, однако они не требуют настройки.

Если построить модель процессов работы модуль прогнозирования результатов успеваемости в общем виде для пользователя модуля (рисунок 3), то получится, что процесс включает в себя этапы запроса пользователем выгрузки из основной системы, автоматического запроса JSON через API. После чего данные сохраняются в БД нейросетевого модуля, и далее обрабатываются нейросетью для получения прогноза. По полученному прогнозу возможно сохранить отчёт [4].

Однако перед внедрением разработанный модуль обязательно протестировать, в том числе — на различных наборах данных, если тестирование пройдено успешно, то можно внедрить нейросетевой модуль в работу высшего учебного заведения.

Самым последним этапом является сбор отчетов об ошибках для мониторинга работы модуля, на основании которых возможно в дальнейшем проводить оптимизацию и добавлять усовершенствования в новых версиях.

С учётом возможной окупаемости разработки и внедрения данного нейросетевого модуля были описаны возможные реализации и нюансы, возникающие при его создании: поскольку по окупаемости может возникнуть вопрос методики ее расчета, — предлагаются следующие формулы окупаемости, где доход ожидается, как следствие повышения инвестиций образовательной организации за счёт успехов обучающихся и внедрения инновационных подходов в образовании:

- P или период, за который окупятся вложения

$$P = \frac{\text{Затраты}}{\text{Доходы}}$$

- NVP или дисконтированный доход

$$NVP = \sum_{t=0}^T \frac{\text{Доходы}_t}{(1+r)^t} - \text{Инвестиции},$$

где T — время, а r — ставка дисконта [5].

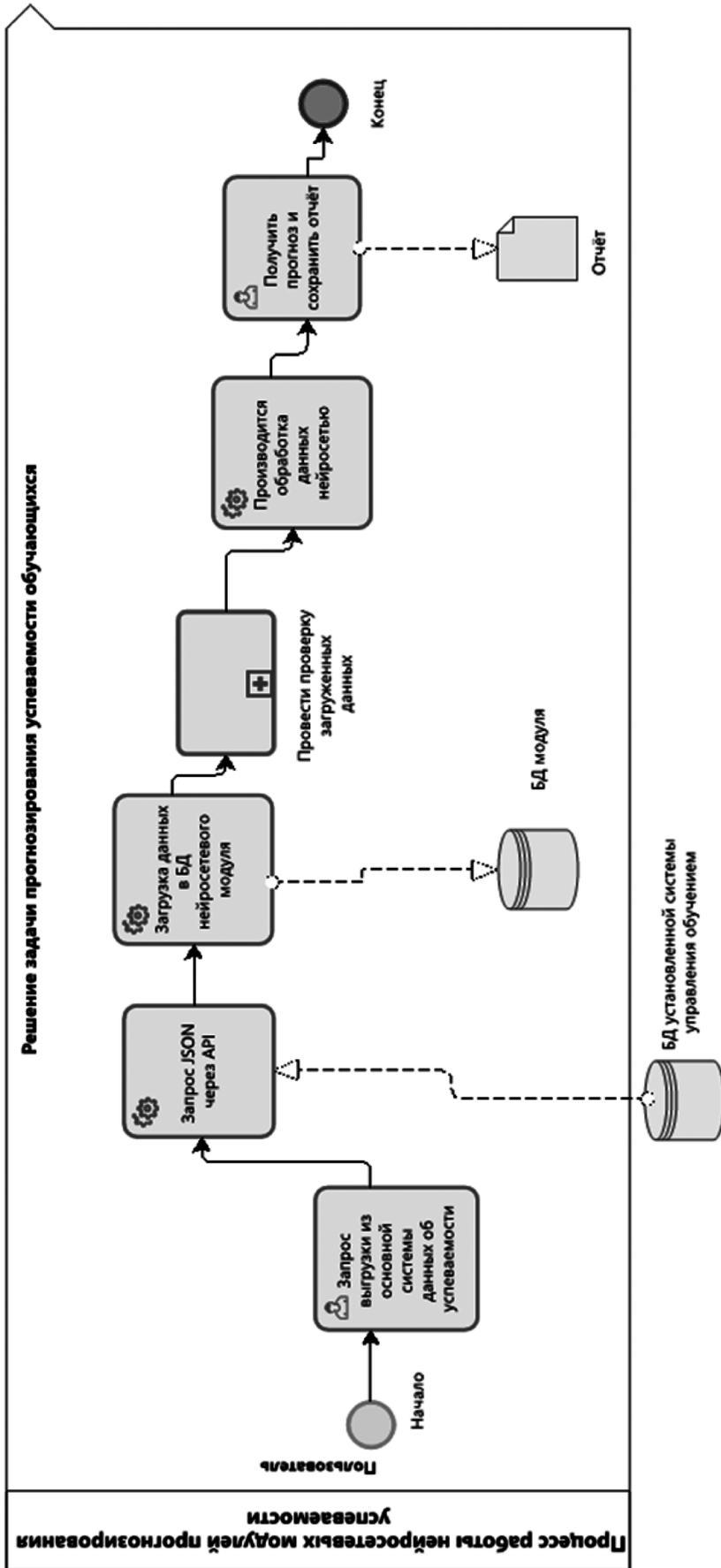


Рис. 3. Процесс работы нейросетевых модулей прогнозирования успеваемости

Заключение

Таким образом, в рамках исследования актуальных вопросов разработки нейросетевых модулей прогнозирования успеваемости были затронуты основные

вопросы, которые могут возникнуть в процессе разработки и внедрения, на основании опыта внедрения подобного программного обеспечения [6].

ЛИТЕРАТУРА

1. Пономарева Л.А., Ромашкова О.Н., Белякова А.Н., Заболотникова В.С. Автоматизация процесса многокритериального ранжирования студентов с помощью электронного портфолио // Вестник Донского государственного технического университета. 2019. Т. 19. № 4. С. 382–388.
2. Белякова А.В., Пономарева Л.А., Чискидов С.В. Прототип информационной системы оценки качества учебного процесса в образовательной организации // В книге: Новые информационные технологии в научных исследованиях. Материалы XXIV Всероссийской научно-технической конференции студентов, молодых ученых и специалистов. — 2019. — С. 45–46.
3. Ponomareva L.A., Chiskidov S.V., Romashkova O.N. Instrumental implementation of the educational process model to improve the rating of the universities // В сборнике: CEUR Workshop Proceedings. 9. Сер. «Selected Papers of the Proceedings of the 9th International Conference Information and Telecommunication Technologies and Mathematical Modeling of High-Tech Systems, ITMM 2019» 2019. С. 92–101.
4. Ponomareva L.A., Romashkova O.N. Training of specialists in on-board communication systems // В сборнике: 2020 Systems of Signals Generating and Processing in the Field of on Board Communications. 2020. С. 9078594.
5. Белякова А.В., Пономарева Л.А., Ромашкова О.Н. Математическая модель оценки качества образовательного процесса // В книге: Новые информационные технологии в научных исследованиях. Материалы XXIV Всероссийской научно-технической конференции студентов, молодых ученых и специалистов. — 2019. — С. 43–44.
6. Белякова А.В., Пономарева Л.А., Чискидов С.В., Василюк И.П. Программа для автоматизированного управления рейтинговыми показателями вузов // Свидетельство о регистрации программы для ЭВМ RU 2019611874, 05.02.2019. Заявка № 2018664941 от 24.12.2018.

© Белякова Анна Вячеславовна (itwhitelight@mail.ru); Ромашкова Оксана Николаевна (ox-rom@yandex.ru);
Ермакова Татьяна Николаевна (ermaktat@bk.ru); Чискидов Сергей Васильевич (chis69@mail.ru)
Журнал «Современная наука: актуальные проблемы теории и практики»