

Догадина Е.П., Осипов А.В.
ФГБОУ ВО Финансовый университет при Правительстве Российской Федерации,
Факультет информационных технологий и анализа больших данных
Департамент анализа данных и машинного обучения
107023, Москва, 4-й Вешняковский пр-д, 4
e-mail: epdogadina@fa.ru, avosipov@fa.ru

Определение коэффициента эффективности процесса выполнения домашних заданий школьниками в условиях пандемии на базе нейронной сети.

Пандемия covid-19 и сопутствующее ей обучение онлайн вызвало большие сложности с освоением учениками школьного материала. Это проявилось в неспособности школьников в заданное время решить установленные по программе задания. Исследовав причины этого явления, было выделено 12 критериев, влияющих на эффективность решения школьных задач. 5 из этих критериев относятся к самой задаче, а 7 – к условиям, при которых она решается. Для определения эффективности решения задач школьниками была использована нейронная сеть MLP (Multilayer Perceptron) реализована с помощью компьютерной программы Neural Excel и обученная методом Resilient Propagation (RProp). Этот метод имеет ряд преимуществ перед другими методами, решающими подобные задачи. Он прост в реализации и имеет высокую скорость сходимости при низких требованиях к погрешности вычисления градиента. Алгоритм использует так называемое «обучение по эпохам», когда коррекция весов происходит после предъявления сети всех примеров из обучающей выборки.

По сравнению с [1] во входной слой нейросети было добавлено дополнительно 5 нейронов, соответствующих пяти параметрам, характеризующим условия решения задачи:

1. Количество затрагиваемых заданием тем
2. Количество видов деятельности, используемых учащимся при выполнении задания
3. Число вопросов в задании
4. Длина задания
5. Сложность формулировки задания
6. Возраст учащегося
7. Пол учащегося
8. Расстояние от дома до школы
9. Средняя оценка по математике
10. Средняя оценка по чтению
11. Режим обучения
12. Доход семьи

• Количество затрагиваемых заданием тем, как величина, определяется т.н. лексемами, характерными для той или иной темы. Например, встречается лексема «сила тока». Это означает, что в задании присутствует тема, связанная с электрическим током.

• Длина задания – количество слов, необходимых для формулировки задания.

• Сложность формулировки задания – экспертная величина, определяемая в баллах в диапазоне от 1 (задание сформулировано просто и однозначно) до 3 (присутствуют избыточные данные и двусмысленные формулировки).

• Возраст учащегося – указывается в годах.

• Пол учащегося – бинарная величина.

• Расстояние от дома до школы – указывается в километрах.

• Средняя оценка по математике за предыдущий период обучения школьника (переводится в пятибалльную систему). В зависимости от школы – это четвертная или триместровая оценка.

• Средняя оценка по чтению для младших классов либо по национальному языку в старших.

• Режим обучения – бинарная величина (онлайн - 0 или офлайн - 1)

- Доход семьи – величина, определяется в баллах от 1 (семья получает дотации от государства) до 3 (семья может позволить себе дорогие машины, проживает в роскошных квартирах и домах и т.д.).

Первые 5 критериев оценивают задачу, как таковую, остальные семь – условия, в которой эта задача решается школьником.

Целевым показателем является число, находящееся в диапазоне от 0 до 1, где 0 – это задание абсолютно бестолковое, совершенно не влияющее на усвоение материала, а 1 – это тот идеал, к которому нужно стремиться (т.е. задание, которое дает возможность учащемуся стопроцентно усвоить материал при максимально комфортных условиях).

Для построения модели использовалась нейронная сеть, состоящая из трех слоев нейронов (рис. 1).

Входной слой – 12 нейронов

Скрытый слой – 19 нейронов

Выходной слой – 1 нейрон.

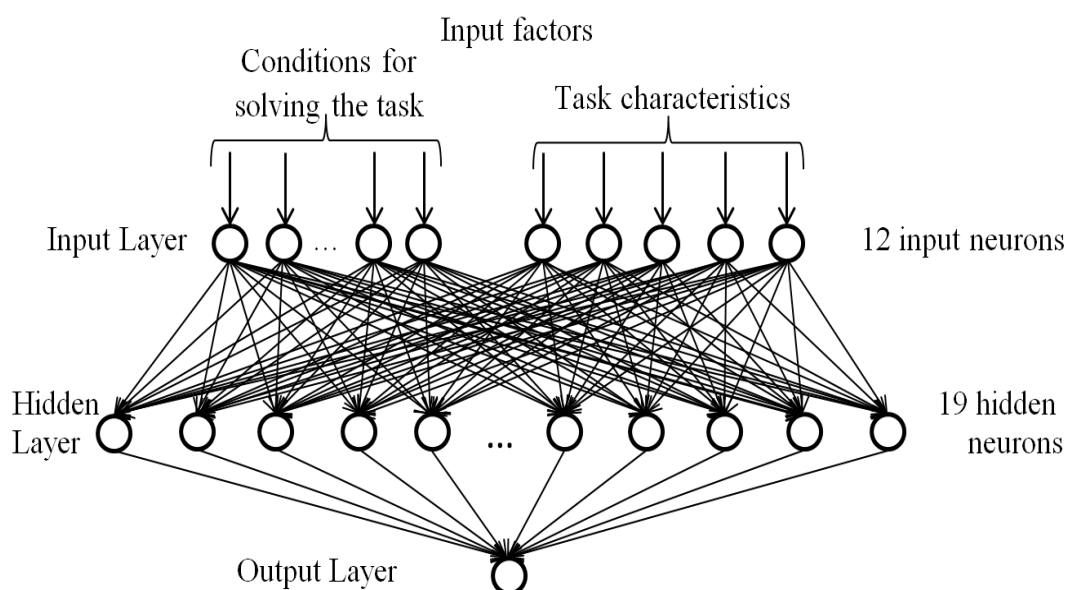


Рис.1. Структура MLP, используемая для определения целевого показателя.

Для обучения нейронной сети было подготовлено обучающее множество, состоящее из 150 заданий по разным предметам и 53 школьников. Приблизительно половина школьников решали эти задачи до пандемии Covid-19, а половина – во время пандемии, для каждого из школьников методом экспертной оценки были выведены значения эффективности каждого задания. Экспертами учитывалась как сама задача, так и условия, влияющие на ее решение. Для тестирования была создана также тестовая выборка из 50 заданий.

На обучающей выборке модель показала 79 % точности прогнозирования эффективности заданий. На тестовом множестве – точность составила 72 %.

Авторами статьи было произведено сравнение результатов, полученных нейросетью при обработке данных до пандемии и во время ее. Было установлено, что средняя эффективность решения задачи каждым школьником во время пандемии уменьшилась на 15%. Кроме того, наблюдается значительное увеличение разброса этого параметра.

Литература

1. Dogadina, E.P.; Smirnov, M.V.; Osipov, A.V.; Suvorov, S.V. Formation of the Optimal Load of High School Students Using a Genetic Algorithm and a Neural Network. Appl. Sci. 2021, Volume 11, Issue 11, 5263, DOI: 10.3390/app11115263.