ИНФОРМАТИКА, ВЫЧИСЛИТЕЛЬНАЯ ТЕХНИКА И УПРАВЛЕНИЕ

УДК 004.032.26

АЛГОРИТМ ОПТИМИЗАЦИИ ГИПЕРПАРАМЕТРОВ МНОГОСЛОЙНОГО ПЕРСЕПТРОНА ДЛЯ РЕШЕНИЯ ЗАДАЧИ ПРОГНОЗИРОВАНИЯ ТЕХНИЧЕСКОГО СОСТОЯНИЯ КОСМИЧЕСКОГО АППАРАТА

Г.А. ЗУБКОВ, научный сотрудник

Главный испытательный космический центр Министерства обороны Российской Федерации им. Г.С. Титова,143090, Краснознаменск, Московская область, ул. Октябрьская, д. 3, e-mail: GeorgyZubkov@mail.ru

© Зубков Г.А., 2021

Статья посвящена разработке алгоритма оптимизации гиперпараметров многослойного персептрона для обучения алгоритма прямого распространения, который решает задачу прогнозирования технического состояния космического аппарата. Актуальность прогнозирования состояния космических аппаратов обусловлена возрастающим уровнем сложности задач, выполняемых сектором анализа центра управления полетом. Предложенный алгоритм включает в себя определение эмпирического распределения значений гиперпараметров со случайным поиском значений из этого распределения, а также проведение дальнейшего поиска значений гиперпараметров модели на сетке. Данный алгоритм позволил существенно сократить время нахождения оптимальных гиперпараметров по сравнению с остальными методами поиска гиперпараметров. Качественная оценка производилась путем сравнения итогового времени нахождения гиперпараметров с использованием алгоритма и применением иных методов.

Ключевые слова: нейронная сеть, многослойный персептрон, гиперпараметры, оптимизация, космический аппарат, машинное обучение, поиск на сетке.

DOI: 10.46573/2658-5030-2021-1-64-70

ВВЕДЕНИЕ

Одной из решаемых задач в проектировании нейронных сетей является настройка и оптимизация гиперпараметров модели. Гиперпараметры, т.е. параметры, устанавливаемые перед обучением нейронной сети, не изменяются в течение всего процесса обучения персептрона.

В различных прикладных пакетах программ гиперпараметры, количество которых может варьироваться от 10 до 50, условно разделены на две категории. К первой относятся гиперпараметры, влияющие на время работы алгоритма и потребление памяти вычислительной машины, ко второй – влияющие на качество получаемой модели и на ее способность максимально точно прогнозировать конечный результат при предъявлении новых примеров. Тем самым возникает противоречие между улучшением одной из категорий параметров и ухудшением другой.

Целью данной работы является разработка алгоритма, позволяющего минимизировать ошибку обобщения модели при ограниченных ресурсах памяти ЭВМ и времени обучения модели.

МАТЕРИАЛЫ И МЕТОДЫ ИССЛЕДОВАНИЯ

В настоящее время существует несколько универсальных методов поиска оптимальных гиперпараметров персептрона: ручная настройка и автоматический поиск, включающий в себя случайный поиск и поиск значений гиперпараметров на сетке значений. Каждый из этих методов имеет свои преимущества и недостатки [1].

Для эффективного ручного поиска следует понимать связи между гиперпараметрами, скоростью обучения, расходом вычислительных ресурсов и ошибкой обучения, следовательно, необходимо понимание фундаментальных основ функционирования нейронных сетей.

Основной задачей обучения многослойного персептрона является минимизация функции ошибки прогнозирования. График обобщения ошибки, как правило, принимает вид U-образной кривой, на одном конце которой значение гиперпараметров соответствует низкой емкости алгоритма и высокой ошибке обобщения, а на другом – высокой емкости алгоритма и высокому значению ошибки обобщения. Данные явления носят названия недообучения и переобучения алгоритма соответственно (рис. 1) [2].

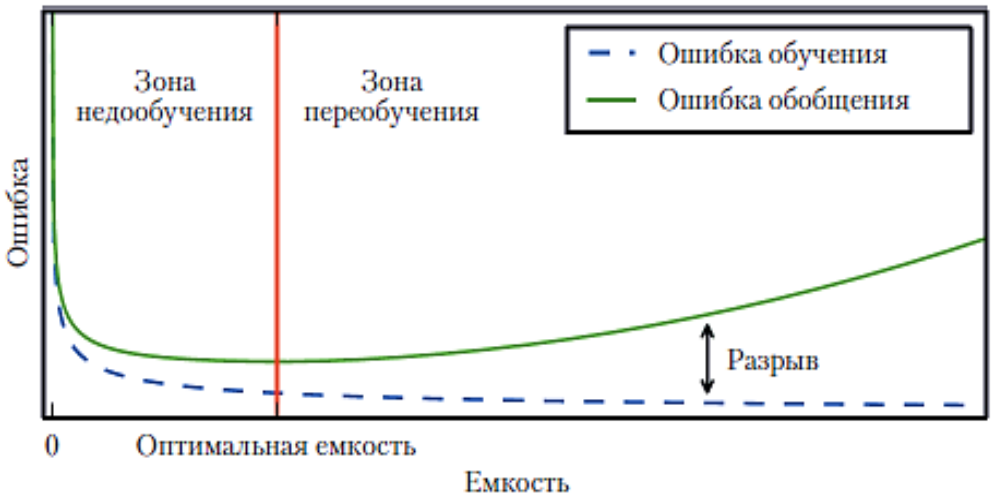

Рис. 1. Зависимость ошибки обучения от эффективной емкости модели [1]

Не все гиперпараметры принимают значения, лежащие на U-образной кривой, часть из них являются бинарными или дискретными, которым соответствует всего несколько точек на U-образной кривой (например, параметры активации, различные переключатели, число блоков в слое и т.д.) [3].

В результате проведенных исследований было установлено, что одним из важнейших гиперпараметров является скорость обучения персептрона, которая оптимальна при максимальной емкости модели. Если скорость обучения слишком велика, то метод градиентного спуска увеличит ошибку обобщения алгоритма; если же скорость обучения слишком мала, то при обучении алгоритм может «застрять» и обучение не будет завершено. Скорость обучения имеет вид U-образной кривой, изображенной на рис. 2 [4].

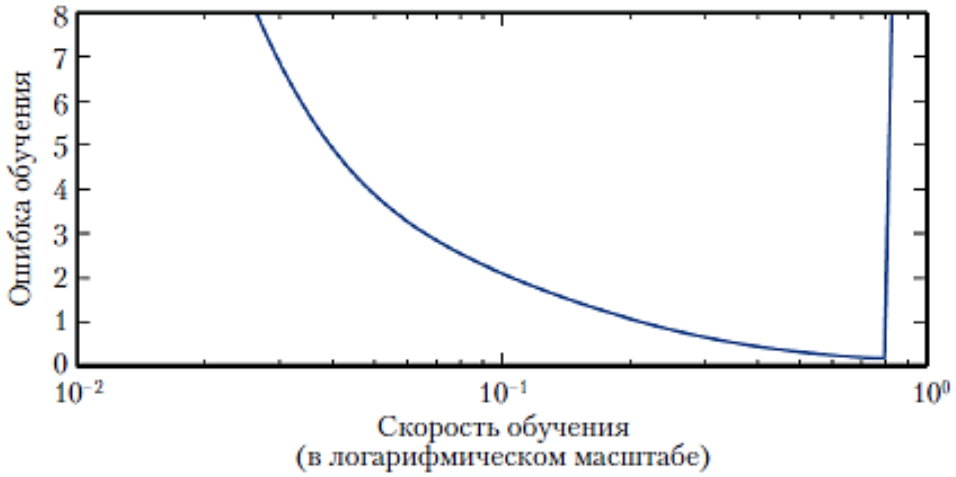

Рис. 2. Зависимость ошибки обучения от скорости обучения [4]

Необходимо помнить, что увеличение или уменьшение определенных параметров ведет к усложнению модели. Эффективная емкость модели ограничена тремя факторами: репрезентативной емкостью модели, способностью к минимизации функции стоимости алгоритма обучения, а также регуляризацией модели. Например, чем больше число слоев модели и число скрытых блоков в каждом слое, тем выше репрезентативная емкость. Однако она не сможет обучиться всем этим функциям, если алгоритм обучения не способен установить, что некоторые функции хорошо минимизируют стоимость, или если регуляризирующие члены, например снижение весов, запрещают использование некоторых функций [5].

В процессе ручной настройки гиперпараметров были обнаружены следующие зависимости:

увеличение числа скрытых блоков и слоев ведет к увеличению времени работы алгоритма и вычислительных мощностей;

оптимальная скорость обучения влечет за собой неизбежное увеличение репрезентативной емкости модели;

увеличение ширины ядра модели слоев приводит к увеличению памяти для хранения параметров и увеличению времени работы алгоритма;

кодировка категориальных переменных *one-hot*-кодированием ведет к увеличению репрезентативной емкости модели;

снижение коэффициентов весов приводит к увеличению репрезентативной емкости модели [6, 7].

Основная рекомендация для ручного поиска связана с необходимостью отслеживания зависимости ошибки обучения и тестирования алгоритма при настройке гиперпараметров. Этого трудоемкого и длительного процесса можно избежать, применяя автоматический поиск гиперпараметров.

Одной из характерных особенностей случайного поиска является отсутствие конкретных значений гиперпараметров, что способствует изучению широких диапазонов значений гиперпараметров без расходования вычислительных ресурсов. Исследование применения случайного поиска показало, что случайный поиск экспоненциально эффективнее поиска на сетке, если присутствуют гиперпараметры, которые не оказывают существенного влияния на качество итоговой модели. Процедуру случайного поиска необходимо проводить несколько раз для уточнения значений, найденных на предыдущих шагах [8].

Если количество изменяемых гиперпараметров модели меньше десяти, часто применяют процедуру поиска значений гиперпараметров на сетке. В данном случае разработчик выбирает конечное множество потенциальных значений гиперпараметров. Как и в процедуре случайного поиска гиперпараметров, лучшими значениями считаются те, при которых значение ошибки обучения персептрона на контрольном наборе данных было минимальным [9].

Одна из важнейших задач в поиске на сетке – правильное определение потенциального множества значений гиперпараметров. Как правило, при поиске на сетке значения выбираются примерно в логарифмическом масштабе. Например, скорость обучения выбирается из множества {0,1; 0,01; 0,001; 0,0001}, число скрытых блоков – в слое {20, 50, 100, 200, 500, 1000, 2000} и так для остальных гиперпараметров. Если же наилучшим значением является экстремум данного массива, то следует задавать иной диапазон для нахождения лучшего значения гиперпараметра [10].

Существенной проблемой процедуры поиска значений гиперпараметров на сетке является экспоненциальный рост вычислительной стоимости при добавлении новых гиперпараметров. Если существует m гиперпараметров, принимающих значение n, то количество экспериментальных сеансов обучения растет как $O(n^m)$. Решением данной проблемы является распараллеливание процессов между серверами, графическими ядрами, а также применение облачных серверов, таких как *Google Colaboratory*.

Рассматриваемый алгоритм состоит из многократного случайного поиска и поиска на сетке значений гиперпараметров. Выполнение данных условий в конечном счете позволит улучшить качество обучения нейронной сети прямого распространения за приемлемое время и ресурсы ЭВМ. Особенностью разработанного алгоритма является сочетание положительных качеств как случайного поиска гиперпараметров, так и поиска на сетке. Предлагаемый алгоритм состоит из следующих шагов.

1. Задаем маргинальное распределение отдельных гиперпараметров. Для категориальных и бинарных параметров задается распределение Бернулли, для остальных гиперпараметров – логнормальное распределение. Распределение гиперпараметра, характеризующего количество скрытых блоков в каждом слое, приведено ниже:

$$\log_number_of_hidden_units \sim u(\log(50), \log(2000)), \quad (1)$$

где $u(\log(50), \log(2000))$ – выборка из равномерного распределения в интервале [50, 2000].

2. Проводим процедуру случайного поиска.

3. Повторяем шаг 2 от 500 до 1 000 раз для проверки предположения и формирования значений, используемых для поиска на сетке.

4. Учитывая ограничения на увеличение репрезентативной емкости модели, формируем вектор из значений гиперпараметров для дальнейшего обучения многослойного персептрона на сетке значений параметров.

5. Проводим процедуру поиска значений гиперпараметров на сетке.

РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ

Применение данного алгоритма позволило существенно сократить временные затраты при поиске оптимального значения гиперпараметров модели на основе многослойного персептрона прямого распространения по сравнению с другими методами (таблица).

Результаты применения различных методов поиска гиперпараметров

Метод поиска	Время поиска гиперпараметров, мин, для повышения точности модели с 67 до 81 %
Ручной поиск значений	63
Случайный поиск	22
Поиск на сетке значений	26
Поиск при помощи алгоритма	11

Результатом проведенного исследования является разработка алгоритма оптимизации гиперпараметров для обучения многослойного персептрона (рис. 3).

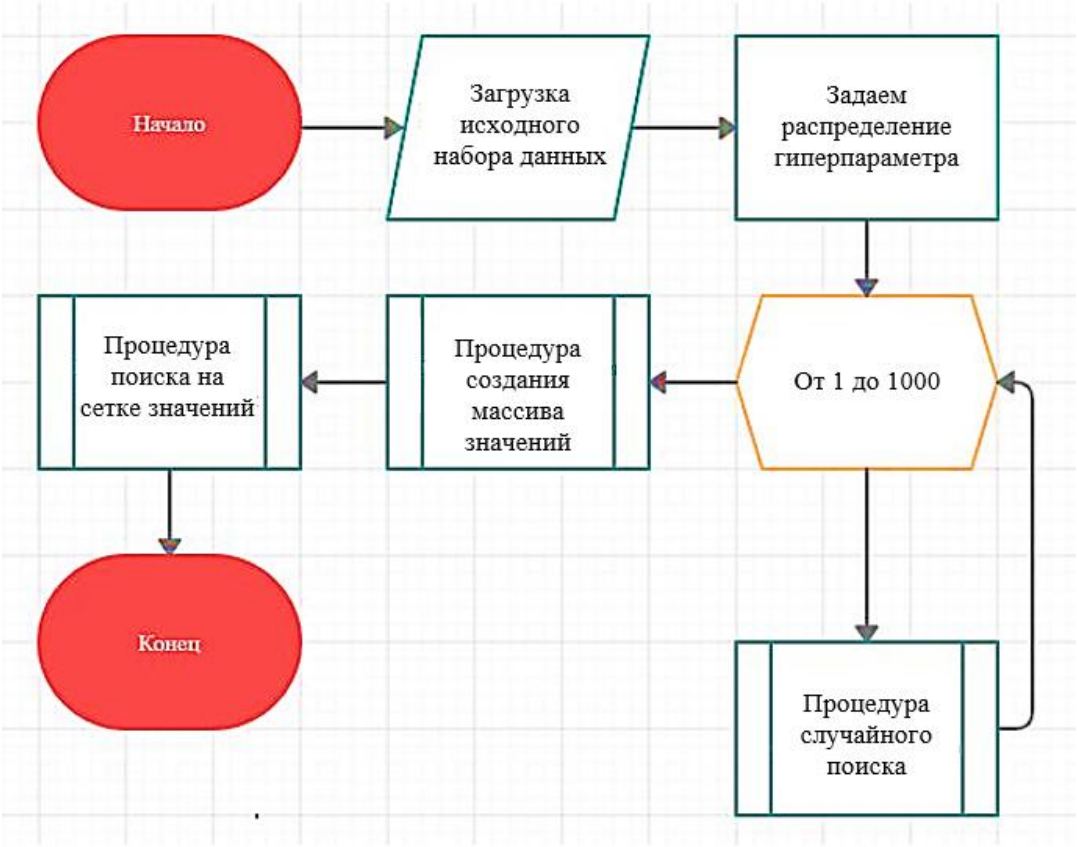

Рис. 3. Алгоритм оптимизации гиперпараметров многослойного персептрона

ЗАКЛЮЧЕНИЕ

Применяя данный алгоритм для оптимизации гиперпараметров многослойного персептрона, удалось ускорить получение гиперпараметров для получения заданной точности прогнозирования технического состояния бортовой аппаратуры космических аппаратов. Полученные результаты должны быть учтены при разработке математической модели интеллектуальной системы анализа телеметрической информации. Данная интеллектуальная система может быть положена в основу разрабатываемого аппаратно-программного комплекса, который позволит в полной мере оптимизировать

процесс анализа статистической информации в секторе анализа центра управления полетом.

ЛИТЕРАТУРА

1. Гудфеллоу Я., Бенджио И., Курвилль А. Глубокое обучение / пер. с англ. А.А. Слинкина. М.: ДМК Пресс, 2018. 652 с.
2. Беркинблит М.Б. Нейронные сети. М.: Моск. институт развития образоват. систем, 1993. 95 с.
3. Ясницкий Л.Н. Введение в искусственный интеллект. М.: Академия, 2005. 176 с.
4. Bergstra J., Bengio Y. Random search for hyper-parameter optimization // *The Journal of Machine Learning Research*. 2012. V. 13. № 1. P. 281–305.
5. Bergstra J.S., Bardenet R., Bengio Y., Kégl B. Algorithms for hyper-parameter optimization // *Advances in neural information processing systems*. 2011. P. 2546–2554.
6. Staudte R.G., Sheather S.J. Robust estimation and testing. New York: Wiley, 1990.
7. Розенблатт Ф. Принципы нейродинамики: перцептроны и теория механизмов мозга. М.: Мир, 1965. 480 с.
8. Раднаев Ч.Б. Оптимизация гиперпараметров сверточной нейронной сети. // Молодежь и современные информационные технологии: сборник трудов XV Международной научно-практической конференции студентов, аспирантов и молодых ученых. Томск: ТПУ, 2017. С. 81–82.
9. Флах П. Машинное обучение. М.: ДМК Пресс, 2015. 400 с.
10. Загоруйко Н.Г. Прикладные методы анализа данных и знаний. Новосибирск: Институт математики, 1999. 268 с.

Для цитирования: Зубков Г.А Алгоритм оптимизации гиперпараметров многослойного персептрона для решения задачи прогнозирования технического состояния космического аппарата // Вестник Тверского государственного технического университета. Серия «Технические науки». 2021. № 1 (9). С. 64–70.

ALGORITHM FOR OPTIMIZING THE HYPERPARAMETERS OF A MULTILAYER PERSEPTRON FOR SOLVING THE PROBLEM OF PREDICTING THE TECHNICAL STATE OF A SPACE VEHICLE

G.A. ZUBKOV, Researcher

Main test space center of the Ministry of Defense of the Russian Federation named after G.S. Titov, 143090, Krasnoznamensk, Moscow region, Oktyabrskaya st. 3, e-mail: GeorgyZubkov@mail.ru

The article is devoted to the development of an algorithm for optimizing the hyperparameters of a multilayer perceptron for training the feedforward algorithm, which solves the problem of predicting the technical state of a spacecraft. The relevance of predicting the state of spacecraft is due to the increasing level of complexity of the tasks performed by the analysis sector of the mission control center. The proposed algorithm includes the determination of the empirical distribution of hyperparameter values with a random search for values from this distribution, as well as further search for the values of the model hyperparameters on the grid. The use of this algorithm made it possible to significantly reduce the time for finding the optimal hyperparameters in comparison with

other methods of searching for hyperparameters. A qualitative assessment was made by comparing the total time of finding the hyperparameters using the algorithm and using other methods.

Keywords: neural network, multilayer perceptron, hyperparameters, optimization, spacecraft, machine learning, grid search.

REFERENCES

1. Goodfellow J., Benjio I., Courville A. Glubokoye obucheniye [Deep learning] / per. s angl. A.A. Slinkina. Moscow: DMK Press, 2018. 652 p.

2. Berkinblit M.B. Neyronnyye seti [Neural networks]. Moscow: Mosk. institut razvitiya obrazovat. sistem, 1993. 95 p.

3. Yasnitsky L.N. Vvedeniye v iskusstvennyy intellekt. [Introduction to artificial intelligence]. Moscow: Akademiya, 2005. 176 p.

4. Bergstra J., Bengio Y. Random search for hyper-parameter optimization. *The Journal of Machine Learning Research*. 2012. V. 13. No. 1, pp. 281–305.

5. Bergstra J.S., Bardenet R., Bengio Y., Kégl B. Algorithms for hyper-parameter optimization. *Advances in neural information processing systems*. 2011, pp. 2546–2554.

6. Staudte R.G., Sheather S.J. Robust estimation and testing. New York: Wiley. 1990.

7. Rozenblatt F. Printsipy neyrodinamiki: pertseptrony i teoriya mekhanizmov mozga [Principles of neurodynamics: perceptrons and the theory of brain mechanisms]. Moscow: Mir, 1965. 480 p.

8. Radnayev Ch.B. Optimization of hyperparameters of a convolutional neural network. *Youth and modern information technologies: collection of works of the XV International scientific-practical conference of students, graduate students and young scientist*. Tomsk: izd-vo TPU, 2017, pp. 81–82. (In Russian).

9. Flakh P. Mashinnoye obucheniye [Machine learning]. Moscow: DMK Press, 2015. 400 p.

10. Zagoruyko N.G. Prikladnyye metody analiza dannykh i znaniy [Applied methods of data and knowledge analysis]. Novosibirsk: Izd-vo instituta matematiki, 1999. 268 p.

Поступила в редакцию/received: 9.10.2020; после рецензирования/revised: 30.10.2020; принята/accepted 27.11.2020