

УДК 519.7

DOI: 10.47813/rosnio.2022.3.131-141

EDN: [SLUSPD](#)



Эволюционный алгоритм проектирования искусственных нейронных сетей с перераспределением ресурсов

П.А. Шерстнев*, Л.В. Липинский

Сибирский государственный университет науки и технологии имени академика М. Ф. Решетнева, пр. им. газ. «Красноярский рабочий», 31, Красноярск, 660037, Россия

*E-mail: sherstpasha99@gmail.com

Аннотация. В статье описывается подход перераспределения ресурсов для обучения искусственных нейронных сетей в процессе эволюции структуры данных сетей. Представлены результаты решения практических задач классификации с применением и без применения предложенного подхода. Показано, что в ряде задач предложенный метод позволяет достичь лучшей точности классификации относительно стандартного подхода при использовании равного количества ресурсов.

Ключевые слова: генетическое программирование, искусственные нейронные сети, генетический алгоритм, классификация

Evolutionary algorithm for designing artificial neural networks with resource redistribution

P.A. Sherstnev*, L.V. Lipinsky

Reshetnev Siberian State University of Science and Technology, 31, Krasnoyarsky Rabochy Av., Krasnoyarsk, 660037, Russia

*E-mail: sherstpasha99@gmail.com

Abstract. The article describes the approach of resource redistribution for training artificial neural networks in the process of evolution of the structure of these networks. The results of solving classification problems with and without the proposed approach are presented. It is shown that in a number of tasks, the proposed method allows achieving better classification accuracy relative to the standard approach when using an equal amount of resources.

Keywords: genetic programming, artificial neural networks, genetic algorithm, classification

1. Введение

Процесс проектирования искусственной нейронной сети (ИНС), как правило, состоит из двух этапов: выбор структуры (количество слоев, нейронов, связей, вид функций активации); настройка весовых коэффициентов связей сети. Для каждой частной задачи требуется особая структура сети, процесс выбора которой является трудоемкой задачей, требующей участия экспертов. Большая часть существующих методик выбора оптимальной структуры разделяются на два класса: задание заведомо большей структуры, чем может быть необходимо для решения задачи, с последующим ее упрощением; создание минимально простой структуры с последующим наращиванием [1]. Данные подходы имеют ряд недостатков, связанных с прямым участием исследователя в процессе построения, а также с ростом числа возможных комбинаций при усложнении сети. Эти проблемы делают актуальным вопрос об разработке и модификации автоматизированных подходов настройки структуры ИНС.

2. Используемый подход

Одним из эффективных подходов автоматизированного проектирования ИНС является подход, предложенный в [2], суть которого заключается в использовании метода генетического программирования (ГП). ГП – это эволюционный метод оптимизации, особенность которого заключается в кодировании решений синтаксическими деревьями, которые в процессе работы эволюционируют, используя принципы эволюционных методов. Внутренними элементами дерева (узлами) являются объекты функционального множества, внешними элементами (листьями) дерева являются объекты терминального множества [3]. В походе [2] функциональное множество состоит из блоков входных и скрытых нейронов. В терминальном множестве находятся две операции: операция объединения боков нейронов в слои (+) и операция последовательного соединения слоев (>). Такой способ кодирования представляет структуру нейронной сети в виде бинарного синтаксического дерева, что позволяет использовать ГП для ее оптимизации.

Для получения оценки эффективности конкретной структуры необходимо сначала обучить ИНС. В данной работе нейросети – индивиды обучались с помощью генетического алгоритма (ГА) со схемой самоадаптации, описанной ниже.

2.1. Метод самоконфигурирования для эволюционных алгоритмов

Ниже описывается подход, предложенный в [4]. Метод основан на увеличении вероятности выбора того оператора, который доставил наибольшую среднюю пригодность на данном поколении. Пусть z_k – число различных операторов k – го типа. Начальная вероятность выбора каждого из операторов определяется как $p_i = 1/z$. Далее, вероятность оператора, доставляющего максимальное среднее значение функции пригодности, увеличивается по формуле (уравнение 1).

$$p_i = p_i + \frac{(z-1)K}{zN}, i = 1, 2, \dots, z \quad (1)$$

Где N – число поколений алгоритма, K – константа, регулирующая скорость изменения вероятности.

Вероятности остальных операторов уменьшаются для равенства вероятности 1. Такой подход можно применять к любым эволюционным методам, таким как ГА и ГП. В данной работе эти алгоритмы используют вышеописанную схему самоадаптации.

2.2. Метод перераспределения ресурсов с использованием алгоритма самоадаптации.

Подход, описанный в [2] большого количества вычислительных ресурсов. Если весовые коэффициенты оптимизируются генетическим алгоритмом, то общее количество вычислений выхода ИНС (количество ресурсов) равно (уравнение 2):

$$total = n_{GA}m_{GA}n_{GP}m_{GP} \quad (2)$$

Где n_{GA} и m_{GA} – количество итераций и размер поколения в ГА, обучающего весовые коэффициенты нейросети, а n_{GP} и m_{GP} – количество итераций и размер поколения в ГП, настраивающего структуру сети.

Задание одинакового размера популяций для ГА на всех итерациях ГП неэффективно. С одной стороны, на первых итерациях структуры сети практически случайные и выделять большое количество ресурсов на случайные структуры может быть не эффективным, с другой стороны на первых итерациях нам требуется не получить готовую нейросеть, а скорее сравнить структуры между собой и отобрать наиболее перспективные. На поздних поколениях в свою очередь необходимо повышать выделяемые ресурсы нейросети, поскольку на них создаются структуры сложнее, чем в начале эволюции. Другая проблема связана с распределением имеющихся ресурсов

между количеством поколений и размером поколения. Можно задавать эти параметры статично (равное количество поколений и размер поколений), но к каждой оптимизационной задаче требуется отдельный подход и в каких-то случаях, например, будет эффективнее повысить количество итераций эволюции в ущерб размеру популяции. Данные проблемы доказывают необходимость в создании эффективной процедуры распределения имеющихся ресурсов в процессе работы алгоритма. Решение первой проблемы может лежать в линейном повышении ресурсов, выделяемых для обучения нейросетей на каждом поколении ГП. Начиная с малого количества ресурсов на ранних итерациях, можно повысить их на поздних. При этом среднее количество будет таким же, если бы они были заданы равными на каждом поколении. Далее, рассмотрим процесс распределение ресурса для обучения каждой нейросети на количество итераций и размер поколений генетического алгоритма. Если N – общее количество вычислений выхода нейросети, то в случае равного распределения можно вычислить (уравнение 3):

$$n_{GA} = m_{GA} = \sqrt{N} \quad (3)$$

Если в данное выражение добавить коэффициент, то можно регулировать распределение ресурсов в пользу количества итераций или размера популяции (уравнение 4, 5):

$$n_{GA} = \sqrt{N}(1 + a) \quad (4)$$

$$m_{GA} = \sqrt{N} \left(\frac{1}{1+a} \right) \quad (5)$$

Где a – коэффициент, регулирующий распределение ресурсов. Данный коэффициент можно изменять в диапазоне от -1 до 1 уменьшая или увеличивая количество итераций в пользу размера популяции.

Далее, для автоматического определения стратегии, можно воспользоваться методом самоадаптации на уровне индивида. В данной работе применяются следующие значения коэффициента a в качестве оператора для самоадаптации (таблица 1).

Таблица 1. Значения коэффициента смещения ресурсов для самоадаптации.

Назначение	Значение a
Смещение ресурсов в сторону итераций	0.5
Равное распределение	0
Смещение ресурсов в сторону популяции	-0.3

Таким образом, используя метод самоадаптации, можно автоматически подбирать эффективное распределение ресурсов между количеством итераций и размером популяции. Данные методики позволяют решать проблемы, описанные в начале параграфа.

При вычислении неизбежно применяется округление до целых чисел. Из-за округления общее количество используемых ресурсов не будет равно стандартному подходу (формула). Предлагается округлять в меньшую сторону, а оставшиеся ресурсы использовать для дообучения нейросети – индивида, полученного по завершении работы алгоритма.

3. Используемые задачи классификации

Оценка эффективности подхода проводилась на основе результатов решения практических задач классификации. Ниже перечислены задачи классификации, используемые в данной работе:

- *Breast Cancer Wisconsin (BCW)*. Диагностика раковых клеток с использованием характеристики ядер клеток [5];
- *Credit risk (CR)* задача принятия решения о выдаче кредита клиенту на основе информации о возрасте, доходе и размере запрашиваемого займа [6];
- *User knowledge modeling data set (UKM)* – задача оценки уровня знаний студента [7];
- *Spam Text Message (STM)*. Тренировочная задача определения спама среди обычных сообщений [8];
- *BBC Full Text Document Classification (BBC)*. Определение тематики сообщения по пяти различным категориям [9];
- *Subjectivity dataset (SD)*. Содержит фрагменты обзоров фильмов из *Rotten Tomatoes*. Задачей является анализ субъективности с целью обозначения мнения как субъективного или объективного [10].

Последние три задачи классификации связаны с обработкой естественных языков. Исходные данные даны в виде текста. Для этих задач признаки были получены с помощью метода построения вложений слов *word2vec* [11]. Размерность вложений составляет 64 признака. Кроме того, для тестирования используется не все доступные данные, а только лишь часть, примерно равная 2 тысячам документов. В таблице ниже

(таблица 2) представлены некоторые характеристики вышеперечисленных наборов данных с учетом обработки текстовых данных.

Таблица 2. Характеристики наборов данных.

Название	Переменных	Всего объектов	Классов
<i>BCW</i>	30	569	2
<i>CR</i>	3	2000	2
<i>UKM</i>	5	403	4
<i>STM</i>	64	1949	2
<i>BBC</i>	64	2225	5
<i>SD</i>	64	1999	2

4. План эксперимента

Цель данной работы доказать эффективность предложенного метода перераспределения ресурсов с использованием алгоритма самоадаптации. Для этого выбранные задачи классификации будут решены двумя альтернативными способами: 1. Для каждой сети – индивида выделяется одинаковое количество ресурсов и их распределение между количеством итераций и размером поколения одинаково (стандартный подход); 2. Количество ресурсов, выделяемых для обучения сетей – индивидов линейно растет с каждым поколением. Количество итераций и размер поколения настраиваются по ходу работы алгоритма с помощью метода самоконфигурации. В дальнейшем, в результатах, этот подход будет маркироваться символом (*L*). Общее количество вычислений выхода нейросети у этих двух подходов равно.

4.1. Гиперпараметры алгоритмов

На основе вычислительных ограничений, а также представленных результатов и параметров в [2] были выбраны следующие гиперпараметры алгоритмов. Для самоконфигурируемого ГП настройки структуры ИНС: количество поколений – 50; размер популяции – 15; размер турнира – 3. для самоконфигурируемого ГА, настраивающего весовые коэффициенты нейросетей: количество ресурсов для обучения сети при равном распределении – 10000 (100 итераций с размером популяции равным 100); количество ресурсов при новом подходе – от 1250 до 18750; число бит на каждую переменную – 12; размер турнира – 3.

4.2. Разделение выборки

Настоящий подход требует разделять данные на три части. Первая часть используется для оптимизации весовых коэффициентов ИНС вида, вторая часть используется, чтобы оценить эффективность структуры и третья часть нужна для того, чтобы оценить эффективность построенной в итоге модели. Для простоты эти части будут называться: обучающая; тестовая; контрольная. Выборки разделены в равных долях. Поскольку природа используемых алгоритмов стохастическая, необходимо оценивать их эффективность не по одиночным запускам, а по многократным. По этой причине, результаты усредняются по 30 – ти запускам, в которых разбиение на подвыборки происходит каждый раз случайно.

4.3. Метрики качества моделей

Для того, чтобы оценить эффективность полученной модели часто достаточно использовать долю правильно классифицированных объектов относительно всех (Точность). Однако, такая метрика имеет ряд недостатков, которые особенно проявляются в задачах, где количество классов не сбалансировано. Поэтому в данной работе используется метрика, называемая $F-1$ мера [12].

5. Результаты

В таблице ниже (таблица 3) представлены усредненные результаты по 30 – ти запускам. В строках указан набор данных, для которого строилась ИНС, а также факт использования предложенного подхода перераспределения ресурсов (L).

Таблица 3. Характеристики наборов данны.

Набор данных	Обучающая + Тестовая выборка ($F-1$ Мера)	Контрольная выборка ($F-1$ Мера)
<i>BCW</i>	0.980904	0.951785
<i>BCW (L)</i>	0.982580	0.959874
<i>CR</i>	0.987786	0.977768
<i>CR (L)</i>	0.990788	0.983890
<i>UKM</i>	0.964153	0.939957
<i>UKM (L)</i>	0.960362	0.950855
<i>SPAM</i>	0.958959	0.936143
<i>SPAM (L)</i>	0.958854	0.940124
<i>BBC</i>	0.904837	0.879505
<i>BBC (L)</i>	0.951659	0.918057
<i>SD</i>	0.882449	0.853111
<i>SD (L)</i>	0.889877	0.856707

Жирным в таблице выше (таблица 3) выделен тот подход из пары (без и с перераспределением ресурсов), который доставил большую точность. Как видно, для всех 6 – ти задач большую точность доставил новый подход перераспределения, предложенный в данной работе. Однако, это может быть лишь результатом случайности. Поэтому, необходимо провести статистическую проверку, чтобы определить, есть ли значимая разница между ошибкой, полученной различными способами обучения. В качестве статистического критерия был выбран критерий Манна-Уитни [13], поскольку он является непараметрическим тестом. Нулевая гипотеза – нет различий между наблюдениями, уровень значимости 0.1. Ниже представлены результаты статистической проверки (дано значение $p - value$) для всех задач классификации (таблица 4).

Таблица 4. Результаты статистической проверки.

Название	$p - value$
<i>BCW</i>	0.307
<i>CR</i>	0.0107
<i>UKM</i>	0.09
<i>SPAM</i>	0.51
<i>BBC</i>	1.43*10⁻⁹
<i>SD</i>	0.29

Согласно результатам из таблицы выше (таблица 4), для ряда задач значение $p - value$ оказалось ниже порогового (выделены жирным), что говорит об отклонении нулевой гипотезы в этих случаях.

Также, интерес могут представлять вероятности использования операторов для самоконфигурируемого ГП. Для каждого из 12 вариантов решений приводить и анализировать график было бы проблематично, поэтому в качестве примера рассмотрим усредненные изменения вероятностей для задачи *BCW (L)* (рисунок 1).

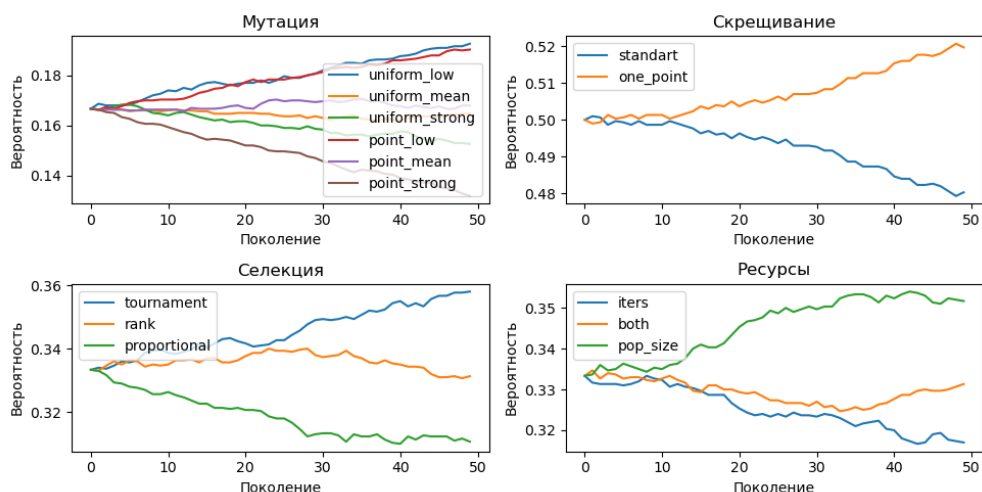


Рисунок 1. Усредненные вероятности использования операторов в ГП.

Дадим некоторые замечания для каждого оператора отдельно. В случае мутации, предпочтительнее использование слабой мутации обоих видов. Операторы скрещивания в течении нескольких начальных поколений соперничают, однако, после этого, доминирующую позицию занимает одноточечная мутация. Из операторов селекции борьба проходила между ранговой и турнирной. Последний, правый нижний график, показывает усредненные вероятности для нового параметра, который определяет способ распределения ресурсов для обучения весовых коэффициентов. На графике видно, что после ряда поколений вперед вырывается стратегия, в которой большая часть ресурсов отдается на размер популяции в генетическом алгоритме.

6. Выводы

В данной работе был описан и протестирован подход эффективного распределения ресурсов для обучения нейронной сети по сравнению со стандартным подходом, в котором на всем протяжении работы алгоритма нейросети используют одинаковые ресурсы. Метод позволяет с тем же суммарным количеством вычислений целевой функции строить сети эффективнее в смысле ошибки классификации. Статистическая проверка показывает, что разница статистически значима для ряда тестовых задач.

Список литературы

1. Никольский, С. М. Линейные уравнения в линейных нормированных пространствах / С.М. Никольский // Изв. АН СССР. Сер. матем. – 1943. – Т. 7, вып. 3. – С. 147-166.

2. Миркес, Е. М. Нейрокомпьютер, проект стандарта / Е. М. Миркес. – Новосибирск "Наука", 1999. – 337 с.
3. Липинский, Л. В. Применение алгоритма генетического программирования в задачах автоматизации проектирования интеллектуальных информационных технологий / Л. В. Липинский, Е. С. Семенкин // Вестник Сибирского государственного аэрокосмического университета им. ак. М.Ф. Решетнева. – 2006. – № 3(10). – С. 22-26.
4. Koza, J. R. Genetic Programming III / J. R. Koza, F. H. Bennet, D. Andre, M. Keane. – Morgan Kaufmann pub., 1999. – ISBN 1-55860-543-6.
5. Семенкина, М. Е. Самоадаптивные эволюционные алгоритмы проектирования информационных технологий интеллектуального анализа данных / М. Е. Семенкина // Искусственный интеллект и принятие решений. – 2013. – №1. – С. 13-23.
6. UCI Machine Learning Repository: Breast Cancer Wisconsin (Diagnostic) Data Set. – Сайт: [Электронный ресурс]. – URL: <https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+%28Diagnostic%29> (дата обращения 17.04.2022).
7. Credit risk | Kaggle. – Сайт: [Электронный ресурс]. – URL: <https://www.kaggle.com/upadorprofzs/credit-risk> (дата обращения 17.04.2022).
8. UCI Machine Learning Repository: User Knowledge Modeling Data Set. – Сайт: [Электронный ресурс]. – URL: <https://archive.ics.uci.edu/ml/datasets/User+Knowledge+Modeling> (дата обращения 17.04.2022).
9. Almeida, T. A. Contributions to the study of SMS spam filtering: New collection and results / T. A. Almeida, J. M. G. Hidalgo, A. Yamakami // DocEng 2011 - Proc. 2011 ACM Symp. Doc. Eng. – 2011. – P. 259-262.
10. Mikolov, T. Efficient Estimation of Word Representations in Vector Space / T. Mikolov, K. Chen, G. Corredo, J. Dean // Proceedings of Workshop at ICLR. Scottsdale, USA. – 2013. – P. 1-12.
11. Pang, Bo. A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts / Bo Pang, Lillian Lee // Proceedings of ACL-04, 42nd Meeting of the Association for Computational Linguistics. – 2004.

12. Mikolov, T. Efficient Estimation of Word Representations in Vector Space / T. Mikolov, K. Chen, G. Corredo, J. Dean // Proceedings of Workshop at ICLR. Scottsdale, USA. – 2013. – P. 1-12.
13. Sasaki, Y. The truth of the F-measure / Y. Sasaki // Teach Tutor mater. – 2007. – pp. 1-5
14. Whitney, H. B. M. On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other / H. B. Mann and D. R. Whitney // The Annals of Mathematical Statistics. – 1947. – № 18(1). – P. 50-60.