ISSN 2523-4714

3. ЛОГИСТИКА

3. LOGISTICS

УДК 004.89

А. М. Туровец, А. Н. Кузьмин
Институт бизнеса БГУ, Минск, Беларусь

ОБОСНОВАНИЕ ЭФФЕКТИВНОСТИ ПРИМЕНЕНИЯ ГЕНЕРАТИВНОГО ИСКУССТВЕННОГО ИНТЕЛЛЕКТА ДЛЯ ОПТИМИЗАЦИИ ЛОГИСТИЧЕСКИХ ПРОЦЕССОВ НА БАЗЕ МЕТОДА НЬЮТОНА

Рассматриваются перспективы применения нейронных сетей для оптимизации различных параметров логистических систем на базе метода Ньютона. Выделены ключевые преимущества и недостатки рассматриваемого метода, которые требуют внедрения новых цифровых решений для повышения общей эффективности. В качестве одного из возможных направлений оптимизации предлагается использовать языковую генеративную модель нейронной сети, которая обучена на матрицах различных порядков. Разработанная модель на базе генеративного искусственного интеллекта существенно сокращает время вычисления обратной матрицы Гессе, что позволяет повысить эффективность оптимизации логистических процессов. Предложенная модель апробирована на результатах сравнения временных затрат на вычисление обратной матрицы Гессе с использованием табличного процессора и нейронной сети. Эмпирические данные позволяют сделать вывод, что частный случай применения нейронной сети может существенно сократить время необходимых вычислений.

Ключевые слова: *нейронные сети, метод Ньютона, языковая генеративная модель, оптимизация логистических процессов, матрицы Гессе*

Для цитирования: Туровец, А. М. Обоснование эффективности применения генеративного искусственного интеллекта для оптимизации логистических процессов на базе метода Ньютона / А. М.Туровец, А. Н. Кузьмин // Бизнес. Инновации. Экономика : сб. науч. ст / Ин-т бизнеса БГУ. – Минск, 2024. – Вып. 9. – С. 158–164.

A. Turovets, A. Kuzmin
School of Business of BSU, Minsk, Belarus

JUSTIFICATION OF EFFICIENCY OF GENERATIVE ARTIFICIAL INTELLIGENCE APPLICATION FOR OPTIMIZATION OF LOGISTIC PROCESSES ON THE BASIS OF NEWTON'S METHOD

This article discusses the prospects of using neural networks to optimize various parameters of logistics systems based on Newton's method. The key advantages and disadvantages of the considered method are highlighted, which require the introduction of new digital solutions to improve overall efficiency. As one of the possible directions of optimization it is proposed to use a language generative model of neural network, which is trained on matrices of different orders. The developed model based on generative artificial intelligence significantly reduces the computation time of the inverse Hesse matrix, which will improve the efficiency of optimization of logistics processes. The proposed model is tested on the results of comparison of time costs for calculating the inverse Hesse matrix using a table processor and a neural network. The empirical data allow us to conclude that a special case of neural network application can significantly reduce the time of necessary calculations.

Keywords: *neural networks, Newton's method, language generative model, optimization of logistic processes, Hesse matrices*

For citation: Turovets A., Kuzmin A. Justification of efficiency of generative artificial intelligence application for optimization of logistic processes on the basis of Newton's method. *Biznes. Innovatsii. Ekonomika = Business. Innovations. Economics.* Minsk, 2024, iss. 9, pp. 158–164.

Введение

В настоящее время в рамках цифровизации экономических процессов продолжается внедрение в деятельность предприятий и организаций инновационных решений, которые призваны повысить эффективность их деятельности на оперативном, тактическом и стратегическом уровнях. Наиболее динамичное развитие на данный момент получают различные области информационных технологий и, в частности, способы машинного обучения, к которым относятся нейронные сети.

Основная часть

Нейронная сеть представляет собой математическую модель, которая базируется на биологических процессах человеческого мозга, происходящих в процессе мышления, а также построена по принципу работы биологических нейронных сетей. Высокий уровень адаптивности к изменяющимся входным данным позволяет получать на последующих этапах обучения более лучшие результаты, чем на предыдущих стадиях. Таким образом, происходит последовательное совершенствование выходных данных (результатов), генерируемых нейронной сетью, на основе изначально заданных входных данных или же изменение конечного результата в соответствии с изменившимися в процессе решения задачи входными условиями.

Для успешной реализации задачи по обучению нейронных сетей и расширения спектра сфер их применения используются математические методы и инструменты для выявления нелинейных зависимостей в данных. Рассмотрим некоторые из них.

Линейная алгебра. Нейронные сети оперируют с векторами, матрицами, тензорами и другими линейными структурами, которые позволяют эффективно хранить, обрабатывать и передавать информацию. Линейная алгебра также используется для вычисления операций, таких как сложение, умножение, транспонирование, обращение, сингулярное разложение и др.

Математический анализ. Нейронные сети используют функции, производные, интегралы, ряды, пределы и другие понятия математического анализа для описания и оптимизации своей структуры и поведения. Математический анализ также применяется для анализа свойств и характеристик нейронных сетей, таких как сходимость, устойчивость, аппроксимация, обобщение и др.

Теория вероятностей и математическая статистика. Нейронные сети используют вероятностные модели, распределения, оценки, гипотезы, тесты, корреляции и другие инструменты теории вероятностей и математической статистики для работы с неопределенностью, шумом, выбросами и другими особенностями данных. Теория вероятностей и математическая статистика также используются для оценки качества и достоверности результатов нейронных сетей, таких как точность, полнота, F-мера, ROC-кривая и др.

Оптимизация. Нейронные сети используют методы оптимизации, такие как градиентный спуск, стохастический градиентный спуск, метод Ньютона, метод сопряженных градиентов, генетические алгоритмы и др., для нахождения оптимальных значений параметров нейронных сетей, например, веса, смещения, скорости обучения, коэффициента затухания и др. Оптимизация также используется для минимизации функции потерь, которая измеряет разницу между желаемым и фактическим выходом нейронной сети.

Метод Ньютона – это один из методов оптимизации, который использует вторую производную (матрицу Гессе) функции потерь для нахождения ее минимума. Метод Ньютона может применяться в нейронных сетях логистики для оптимизации параметров нейронных сетей (веса и смещения). Матрица Гессе – это квадратная симметричная матрица вторых частных производных функции $f(\times 1, ..., \times n)$. Геометрически эта матрица описывает изменение градиента функции. Она позволяет исследовать поведение функции в окрестности точки. Матрица Гессе используется в задачах оптимизации, таких как поиск экстремумов функции или решение систем нелинейных уравнений.

В логистике матрица Гессе может применяться для анализа и оптимизации различных логистических процессов, например планирование маршрутов, распределение ресурсов, управление запасами, формирование заказов, ценообразование и др. Например, матрица Гессе может помочь в определении наилучшего способа доставки товаров от поставщика к потребителю, учитывая стоимость, время, расстояние, спрос, предложение и другие факторы. Для вычисления матрицы Гессе можно использовать различные методы, такие как аналитический, численный, квазиньютоновский и др. В зависимости от сложности функции и размера матрицы эти методы могут иметь разную вычислительную сложность, точность и устойчивость.

Предположим, что мы хотим оптимизировать распределение грузовых автомобилей между несколькими складами, чтобы минимизировать общие транспортные издержки [1]. Для этого можно использовать следующую функцию потерь:

$$L(x)=\sum_{i=1}^{n}\sum_{j=1}^{m}c_{ij}x_{ij}$$

где n – количество складов; m – количество автомобилей; c_{ijj} – стоимость доставки автомобиля j от склада i; x_{ij} – бинарная переменная, равная 1, если автомобиль j доставляется от склада i, и 0 в противном случае.

Кроме того, мы имеем следующие ограничения:

$$\sum_{i=1}^{n}x_{ij}=1,\ \forall j=1,\ \dots,\ m,$$

$$\sum_{i=1}^{m}x_{ij}\le s_i,\ \forall j=1,\ \dots,\ n,$$

где s_i – максимальное количество автомобилей, которые может отправить склад i.

Для решения этой задачи мы можем использовать метод Ньютона, который требует вычисления матрицы Гессе для функции потерь.

Матрица Гессе имеет размер $(nm)\times(nm)$ и состоит из следующих блоков:

$$H_{ij}=\frac{d^2L}{dx_idx_j}=\begin{cases} c_{ij} \text{ если } i=j, \\ 0, \text{ если } i\ne j. \end{cases}$$

Таким образом, матрица Гессе имеет диагональный вид и равна:

$$H=\begin{bmatrix} c_{11} & 0 & \dots & 0 \\ 0 & c_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & c_{nm} \end{bmatrix}.$$

Зная матрицу Гессе, мы можем применить метод Ньютона, который на каждом шаге обновляет решение x по формуле

$$x^{(k+1)}=x^{(k)}-H^{-1}\forall L\left(x^{(k)}\right).$$

где $\forall L\left(x^{(k)}\right)$ – градиент функции потерь, вычисленный в точке $x^{(k)}$, а H^{-1} – обратная матрица Гессе. При этом мы должны учитывать ограничения на переменные x_{ij} и проверять условие остановки, например, достижение заданной точности или максимального числа итераций.

Однако метод Ньютона имеет ряд следующих недостатков:

– высокая сложность вычислений в связи с операциями над матрицей Гессе на каждом шаге, что может вызвать затруднения в работе нейронной сети при больших размерах системы;

– неустойчивость, потому что метод Ньютона может сходиться к локальному минимуму или седловой точке, а не к глобальному минимуму функции потерь;

– чувствительность к выбору начальной точки, поскольку метод Ньютона может давать разные результаты в зависимости от того, с какой точки начинается оптимизация [2].

На практике сегодня чаще встречаются другие методы оптимизации: градиентный спуск, стохастический градиентный спуск, метод сопряженных градиентов и другие, которые имеют меньшую вычислительную сложность и устойчивость которых больше, однако метод Ньютона может найти применение в логистических системах различной сложности при условии увеличения его скорости за счет генеративного искусственного интеллекта, что, в свою очередь, позволит существенно сократить расходы на его проведение.

В основе определения трудоемкости нахождения обратной матрицы Гессе лежит утверждение, что она изменяется прямо пропорционально изменению количества задействованных для оптимизации элементов. Также увеличение трудоемкости нахождения обратной матрицы Гессе зависит и от выбранного алгоритма, используемого для этой задачи. Существуют разные методы, такие как метод алгебраических дополнений, метод Гаусса – Жордана, метод разложения и др. Каждый из них имеет свою сложность, которая обычно выражается в виде функции от порядка матрицы n.

Например, метод алгебраических дополнений имеет сложность $O(n^5)$, т. е. для нахождения обратной матрицы нужно выполнить порядка n^5 элементарных операций. Это означает, что при увеличении порядка матрицы в два раза, трудоемкость возрастает в $2^5 = 32$ раза. Этот метод считается неэффективным для больших матриц.

Метод Гаусса – Жордана имеет сложность $O(n^3)$, т. е. для нахождения обратной матрицы нужно выполнить порядка n^3 элементарных операций. Это означает, что при увеличении порядка матрицы в два раза, трудоемкость возрастает в $2^3 = 8$ раз. Этот метод считается более эффективным, чем предыдущий, но все еще требует много времени для больших матриц.

Метод разложения имеет сложность $O(n^3)$, но с меньшим коэффициентом, чем метод Гаусса – Жордана. Этот метод основан на представлении матрицы в виде произведения двух или трех матриц, для которых легко найти обратные. Например, можно использовать разложение видов LU, QR или SVD. Этот метод считается наиболее эффективным для нахождения обратной матрицы, но требует дополнительной памяти для хранения разложения [3].

В связи с тем, что метод разложения имеет максимальную эффективность при нахождении обратной матрицы Гессе с увеличенным порядком, целесообразно внедрить языковую генеративную модель нейронной сети, которая генерирует выходные данные в рамках ограничения в виде поставленной задачи – нахождения обратной матрицы Гессе методом разложения при ограничениях, заданных начальным условием задачи.

Как уже было отмечено, процесс нахождения обратной матрицы Гессе методом разложения включает в себя представление такой матрицы, как произведение двух или же трех матриц, для которых найти обратные легче. На этой основе составим таблицу трудоемкости нахождения обратной матрицы для матриц второго и третьего порядков (табл. 1).

Таблица 1

Нахождение обратной матрицы для матриц второго порядка

Table 1

Findingtheinversematrixforsecond-ordermatrices

Действие	Алгоритм
Нахождение определителя матрицы	$\lvert A\rvert = A_{11} \times A_{22} - A_{12} \times A_{21}$
Нахождение транспонированной матрицы алгебраических дополнений	После составления матрицы миноров второго порядка, необходимо изменить знаки элементов второй диагонали, получив матрицу алгебраических дополнений. Далее выполняется транспонирование (изменение расположения строк и столбцов исходной матрицы так, чтобы строки стали столбцами, а столбцы – строками).
Вычисление обратной матрицы	$A^{-1} = \frac{1}{\lvert A\rvert} \times A_{ij}^{T}$

И с т о ч н и к: разработано автором.
S o u r c e: author’s developed.

В данном случае под трудоемкостью будет пониматься необходимое количество итераций для нахождения матриц нужного порядка, независимо от их размерности. Для характеристики следующего этапа составляем таблицу трудоемкости нахождения обратной матрицы для матриц третьего порядка (табл. 2).

Таблица 2

Нахождение обратной матрицы для матриц третьего порядка

Table 2

Findingtheinversematrixforthird-ordermatrices

Действие	Алгоритм
Нахождение определителя матрицы	$\lvert A \rvert = A_{11} \times A_{22} \times A_{33} + A_{21} \times A_{32} \times A_{13} + A_{12} \times A_{23} \times A_{31} - -A_{31} \times A_{22} \times A_{13} - A_{21} \times A_{12} \times A_{33} - A_{23} \times A_{31} \times A_{11}$
Нахождение транспонированной матрицы алгебраических дополнений	После составления матрицы миноров третьего порядка необходимо изменить знаки в элементах A_{12}, A_{21}, A_{23} и A_{32}, получив матрицу алгебраических дополнений A_{ij}^T. Далее переходим к выполнению процедуры транспонирования
Вычисление обратной матрицы	$A^{-1} = \frac{1}{\lvert A \rvert} \times A_{ij}^T$

И с т о ч н и к: разработано автором.
S o u r c e: author's developed.

Как видно из табл. 1 и 2, трудоемкость нахождения обратной матрицы Гессе с увеличением порядка матрицы возрастает. Такое повышение трудоемкости выражается в существенных временных затратах на поиск лишь одного компонента формулы метода Ньютона – обратной матрицы порядка *n*. При увеличении порядка матрицы в два раза трудоемкость возрастает в 8 раз.

В связи с этим предлагается использовать языковую генеративную нейронную сеть с ограничениями по выходным данным в виде поставленной задачи – найти обратную матрицу Гессе методом разложения, основываясь на начальном условии задачи. Ограничения спектра выполняемых задач будут заданы в начале работы и сохранены в памяти нейронной сети. Они будут рассматриваться ей как долговременная зависимость, которая не изменится до отмены его пользователем. Для ускорения вычислительных операций нейронной сети следует также выполнить первичное ее обучение путем загрузки массива результатов вычисления обратных матриц Гессе для матриц второго и третьего порядков. На основе результатов обучения матрицы большего порядка также будут вычисляться.

Средние временные затраты на нахождение обратной матрицы третьего порядка при использовании табличного процессора, включая построение самой матрицы плюс проверку, даны в табл 3.

Таблица 3

Средние временные затраты на нахождение обратной матрицы третьего порядка при использовании табличного процессора с проверкой

Table 3

The average time spent on finding the inverse matrix of the third order when using a table processor with validation

Действие	Средние временные затраты, с
Запуск табличного процессора	7
Перенос матрицы порядка *n* в табличный процессор (при наличии матрицы Гессе)	10
Ввод формулы в табличный процессор для вычисления элементов обратной матрицы порядка *n*	8
Проверка правильности построения обратной матрицы Гессе	23
Итого:	48

И с т о ч н и к: разработано автором.
S o u r c e: author's developed.

В итоге средние временные затраты на эту операцию с использованием табличного процессора составляют 48 с. Временные затраты можно сократить при помощи внедрения нейронной сети, предварительно обученной на матрицах второго и третьего порядков. В качестве подтверждения сокращения временных затрат приведем время вычисления обратной матрицы Гессе третьего порядка, предварительно обученной языковой генеративной моделью нейронной сети (табл. 4).

Таблица 4

Временные затраты на нахождение обратной матрицы порядка *n* с помощью нейронной сети, обученной на матрицах второго и третьего порядков

Table 4

The time spent on finding the inverse matrix of order n using a neural network trained on second- and third-order matrices

Измерение, n	1	2	3	4	5	6	7	8	9	10
Время, с	8,4	7,5	9,0	6,6	9,1	7,2	7,0	8,0	7,1	6,5

И с т о ч н и к: разработано автором.
S o u r c e: author's developed.

С учетом данных табл. 4 выяснено, что среднее значение временных затрат в процессе построения нейронной сетью обратной матрицы Гессе порядка *n* составляет:

$$\frac{8{,}4+7{,}5+9{,}0+6{,}6+9{,}1+7{,}2+7{,}0+8{,}0+7{,}1+6{,}5}{10}=7{,}64 \text{ с.}$$

Сравнив временные затраты на данную операцию с использованием табличного процессора с временными затратами с использованием нейронной сети, предварительно обученной на матрицах второго и третьего порядков, можно заметить, что во втором случае экономия временных затрат существенна:

$$48 \text{ с} / 7{,}64 \text{ с} = 6{,}28 \text{ раз.}$$

Таким образом, выделенное время на совершение одной такой операции сократится с 48 с до 7,64 с, что означает снижение времени выполнения одной операции на 40,36 с, при условии сравнения средних временных затрат на выполнение данной операции.

Сама схема работы нейронной сети – следующая: перед внедрением нейронной сети необходимо составить шаблон типовой задачи, который мог бы заполняться необходимыми начальными условиями в автоматическом режиме путем извлечения начальных условий из базы данных. Затем оператор (пользователь) вводит шаблон задачи по нахождению обратной матрицы Гессе с учетом ограничений, далее нейронная сеть вычисляет обратную матрицу Гессе порядка *n*, а оператор повторяет эту операцию после предыдущей операции, сообщая нейросети об изменениях во входных условиях (например, в ограничениях, числовых значениях матрицы Гессе и др.). Получив обратную матрицу Гессе от нейросети, можно перейти к исполнению метода Ньютона.

Языковую генеративную модель нейронной сети предлагается распространить не только на эту часть метода Ньютона, но и на всю формулу, которая вычисляет и обращает матрицу Гессе на каждом шаге. Однако внедрение нейронной сети на весь метод Ньютона предлагается производить последовательно ввиду того, что генеративные языковые модели нейронных сетей в процессе своего обучения могут совершать существенные или же несущественные ошибки, что в конечном итоге все равно дает неправильный результат. В процессе внедрения необходимо наладить процесс постоянного обновления исходных данных путем включения элементов в уже существующую систему, например, новых складов. Вся работа нейронной сети в долгосрочной перспективе с учетом изменения параметров системы также будет влиять на итоговый результат, находя новые минимальные значения.

Заключение

Таким образом, для решения различных оптимизационных задач в рамках логистических систем, предлагается использовать языковую генеративную модель нейронной сети, которая обучена на матрицах второго и третьего порядков. Данная модель позволяет существенно сократить время вычисления обратной матрицы Гессе, что приводит к увеличению эффективности оптимизации логистических процессов, обеспечивая высокую точность и вычислительную скорость.

Список использованных источников

1. *Туровец, А. М.* Поиск альтернативных вариантов доставки грузов в международном сообщении с учетом требований санкционных ограничений // А. М.Туровец, Е. П. Выдрицкая, А. Е. Якушева // Автомобилестроение: проектирование, конструирование, расчет и технологии ремонта и производства : материалы VII Всероссийской науч.-практ. конф., Ижевск, 28–29 апреля 2023 г. / под ред. Н. М. Филькина. – Ижевск : УИР ИжГТУ имени М. Т. Калашникова, 2023. – С. 591–596

2. *Chen, H.* Topographic property of back propagation artificial neural network: From human functional connectivity network to artificial neural network / H. Chen, F. Lu, B. He // Neurocomputing. – 2020. – P. 200 – 210.

3. *Бондарев, П. А.* Обучение нейросети на базе шарового метода оптимизации Ньютона [Электронный ресурс] / П. А. Бондарев, Р. А. Проскурин // Наука и образование: науч. изд. МГТУ им. Н. Э. Баумана. – 2005. – № 6. – Режим доступа: http://engineering-science.ru/doc/51953.html. – Дата доступа: 14.03.2024.

References

1. Turovec A. M., Vydrickaya E. P., Yakusheva A. E. Search for alternative options for the delivery of goods in international traffic, taking into account the requirements of sanctions restrictions. *Avtomobilestroenie: proektirovanie, konstruirovanie, raschet i tekhnologii remonta i proizvodstva : materialy VII Vserossijskoj nauchno-prakticheskoj konferencii* [Automotive industry: design, construction, calculation and repair and production technologies : materials of the VII All-Russian Scientific and Practical Conference]. Izhevsk, 2023, pp. 591– 596 (in Russian).

2. Chen H., Lu F., He B. Topographic property of back propagation artificial neural network: From human functional connectivity network to artificial neural network. *Neurocomputing*, 2020, pp. 200–210.

3. Bondarev P. A., Proskurin R. A. Neural network training based on Newton's spherical optimization method. *Nauka i obrazovanie: nauchnoe izdanie MGTU im. N. E. Baumana = Science and Education: scientific edition of Bauman MSTU.* 2005, no 6. Available at: http://engineering-science.ru/doc/51953.html(accessed 14 March 2024) (in Russian).

Информация об авторах

Туровец Александр Михайлович – старший преподаватель кафедры логистики, Институт бизнеса БГУ, e-mail: imprudance@gmail.com

Кузьмин Артём Николаевич – студент 3 курса, Институт бизнеса БГУ, e-mail: artem160104k@gmail.com

Information about the authors

Turovets, A. – Senior lecturer at the Department of logistics, School of Business of BSU, e-mail: imprudance@gmail.com

Kuzmin, A. – 3rd year student; School of Business of BSU, e-mail: artem160104k@gmail.com

Статья поступила в редколлегию 30.04.2024
Received by editorial board 30.04.2024