

Научная статья

Статья в открытом доступе

УДК 004.8

doi: 10.30987/2658-6436-2023-2-50-57

УПРАВЛЕНИЕ КАЧЕСТВОМ МОДЕЛЕЙ РЕГРЕССИИ НА ОСНОВЕ ЗАДАЧ МАТЕМАТИЧЕСКОГО ПРОГРАММИРОВАНИЯ

Павел Федорович Чернавин^{1✉}, Николай Павлович Чернавин², Федор Павлович Чернавин³

^{1, 2, 3} Уральский федеральный университет им. первого Президента России Б.Н. Ельцина, г. Екатеринбург, Россия

¹ chernavin.p.f@gmail.com, <http://orcid.org/0000-0003-3214-3906>

² ch_k@mail.ru, <http://orcid.org/0000-0003-3214-3906>

³ chernavin_fedor@mail.ru, <http://orcid.org/000-0003-3214-3906>

Аннотация. *Качество моделей регрессии необходимо оценивать многими показателями. Критериями качества могут быть минимум суммы квадратов или абсолютных величин отклонений прогнозируемых значений от истинных, адекватность величины и знака коэффициентов в уравнениях регрессии, робастность модели, минимум признаков необходимых для выполнения других показателей и многое другое. При построении уравнений регрессии стандартными программами достаточно сложно одновременно учесть несколько из перечисленных показателей. Цель статьи – продемонстрировать, что построение моделей регрессии на основе задач математического программирования позволяет учитывать в рамках одной модели большую совокупность требований к качеству решения одновременно. Научная новизна состоит в том, что такой подход позволяет создавать более сложные модели регрессии, учитывающие специфику конкретных практических задач. Например, в генеральной выборке могут одновременно присутствовать различные тенденции. В этом случае необходимо выяснить сколько уравнений регрессии требуется для описания имеющихся наблюдений с заданной точностью. Частным случаем такой постановки является кусочно-линейная регрессия. Другим примером может быть необходимость прогнозирования нескольких выходных параметров минимальным набором одинаковых входных параметров. В статье приведены практические результаты применения авторского подхода для решения задач регрессии в агломерационном производстве и прогнозировании финансовых результатов для банковской сферы.*

Ключевые слова: математическое программирование, минимизация количества признаков, регрессии: линейная, нелинейная, лассо, гребневая, эластичная, кусочно-линейная

Для цитирования: Чернавин П.Ф., Чернавин Н.П., Чернавин Ф.П. Управление качеством моделей регрессии на основе задач математического программирования // Автоматизация и моделирование в проектировании и управлении. 2023. №2 (20). С. 50-57. doi: 10.30987/2658-6436-2023-2-50-57.

Original article

Open Access Article

QUALITY CONTROL OF REGRESSION MODELS BASED ON MATHEMATICAL PROGRAMMING PROBLEMS

Pavel F. Chernavin^{1✉}, Nikolai P. Chernavin², Fedor P. Chernavin³

^{1 2 3} Ural Federal University, Yekaterinburg, Russia

¹ chernavin.p.f@gmail.com, <http://orcid.org/0000-0003-3214-3906>

² ch_k@mail.ru, <http://orcid.org/0000-0003-3214-3906>

³ chernavin_fedor@mail.ru, <http://orcid.org/000-0003-3214-3906>

Abstract. *The quality of regression models must be evaluated by many indicators. Quality criteria can be the minimum of square sum or absolute values of deviations of the predicted values from the true ones, the adequacy of value and sign of the coefficients in the regression equations, the model robustness, the minimum of signs necessary to fulfil other indicators, and much more. When constructing regression equations using standard programmes, it is quite difficult to simultaneously take into account several of the listed indicators. The aim of the article is to demonstrate that building regression models based on mathematical programming problems allows simultaneously considering a large set of requirements for the solution quality within one model. The scientific novelty lies in the fact that this approach*

makes it possible to create more complex regression models that take into account the specifics of particular practical problems. For example, in the general sample, there may be different trends at the same time. In this case, it is necessary to find out how many regression equations are required to describe the available observations with a given accuracy. A special case of such a formulation is piecewise linear regression. Another example can be the need to predict multiple output parameters with a minimal set of identical input parameters. The article presents the practical results of applying the author's approach to solving regression problems in agglomeration production and forecasting financial results for the banking sector.

Keywords: mathematical programming, regressions: linear, nonlinear, lasso, ridge, elastic, piecewise linear; minimization of the number of features

For citation: Chernavin P.F., Chernavin N.P., Chernavin F.P. Quality control of regression models based on mathematical programming problems. Automation and modeling in design and management, 2023, no. 2(20). pp. 50-57. doi: 10.30987/2658-6436-2023-2-50-57.

Введение

Несмотря на то, что различные подходы к решению задач регрессии развиваются с времен Гаусса и Лежандра, т.е. с конца 18 века, до сих пор остается большое количество проблем при оценке качества как самих моделей регрессии, так и результатов, полученных на их основе. Данные проблемы достаточно хорошо освещены в работах различных авторов [1 – 3]. На наш взгляд, качество моделей регрессии нельзя сводить только к различным метрикам (среднеквадратичная ошибка, средняя абсолютная ошибка и т.п). Метрики, конечно, важны, но более важным является адекватность всей модели в целом конкретной практической ситуации. Кроме хороших метрик модель должна быть робастной и устойчивой, содержать по возможности минимальное число признаков, учитывать мнение специалистов о величине и знаках коэффициентов в уравнениях регрессии и другие пожелания практических специалистов, которые являются заказчиками конкретного исследования. То есть понятие качества модели имеет много составляющих и всеми ими надо уметь управлять в рамках самой модели. По нашему мнению, многие такие пожелания могут быть учтены, если модели регрессии представлять как задачи математического программирования.

Следует отметить, что именно появление проблем и необходимость учета пожеланий практических специалистов приводит к появлению новых моделей регрессии. Проблемы с мультиколлинеарностью признаков, необходимостью минимизации числа признаков, абсурдностью, в ряде случаев, величины коэффициентов в уравнениях линейной регрессии (ЛР) способствовали появлению моделей Lasso, Ridge и ElasticNet [4 – 5]. Недостатки классических моделей ЛР привели к появлению принципиально новых видов регрессии, основанных на нейросетях, деревьях решений, методе ближайших соседей и других подходах, заимствованных из других направлений машинного обучения (МО). Новые методы, в ряде случаев, способны давать более точные прогнозы, но явно проигрывают модификациям классической ЛР по интерпретируемости решений, так как не дают решения в аналитическом виде, т.е. в виде некоторого уравнения.

Вместе с тем, в качестве функций для построения уравнений регрессии можно использовать и нелинейные, желательно сепарабельные функции. Сепарабельные функции линейны относительно коэффициентов при их составляющих. Поэтому переход в пространство признаков более высокой размерности приведет опять к задаче ЛР. Например, вместо линейной функции будем использовать полином второй степени, тогда вместо задачи:

$$\min \sum_{j \in J} \left(\sum_{i \in I} X_{ij} \cdot a_i + b - Y_j \right)^2 \quad (1)$$

Необходимо будет решить задачу:

$$\min \sum_{j \in J} \left(\sum_{i \in I_1} X_{ij} \cdot a_i + \sum_{i \in I_2} X_{ij}^2 \cdot a_i + b - Y_j \right)^2 \quad (2)$$

где a_i – коэффициенты уравнения регрессии (искомые переменные); b – свободный член (искомая переменная); X_{ij} – входной i -ый признак j -го наблюдения (константы); Y_j – выходной признак j -го наблюдения (константы); I_1 – множество исходных входных параметров;

I_2 – множество сгенерированных входных параметров (квадраты исходных); J – множество наблюдений; i, j – индексы соответствующих множеств.

Очевидно, что задача (2) – это тоже задача обыкновенной ЛР с критерием MSE (Mean Squared Error). Естественно, что вместо параболы можно использовать любой полином или сепарабельную функцию. Поэтому далее возможность нелинейности уравнений регрессии отдельно оговаривать не будем, только заметим, что применение не сепарабельных функций может приводить к абсурдным результатам. Заметим также, что с точки зрения математического программирования (МП) модель (1) – это задача квадратичного программирования, записанная в компактной форме. Для решения таких задач мы использовали программный продукт IBM ILOG CPLEX. Данный пакет упомянут неслучайно, так как он позволяет решать, как задачи квадратичного программирования, так и задачи с целочисленными переменными.

Некоторые авторы в своих публикациях высказывают мнение, что «линейная регрессия часто неприменима к реальным данным из-за слишком ограниченных возможностей и ограниченной свободы маневра» [2]. Предлагаем взглянуть на задачи регрессии более широко как на задачи МП. Для тех, кто владеет этим инструментарием, свобода маневра и возможности учета практически любых пожеланий заказчиков конкретного исследования, на наш взгляд, безграничны. Далее приведем только некоторые пожелания заказчиков, которые встретились нам на практике в совершенно различных предметных областях: финансы, медицина, неразрушающий контроль, металлургия. Отметим, что учет этих пожеланий почти всегда приводил к совершенно новым моделям регрессии, которые удавалось формализовать как задачи МП. Поэтому вначале приведем классические модели в виде задач МП, а затем ряд оригинальных постановок, которые встретились нам на практике.

Классические модификации моделей линейной регрессии в виде задач МП

В принципе, различные классические модификации моделей ЛР отличаются только функцией цели и возникают из желания уменьшить влияние различных факторов на качество решения. Критерий MSE всем хорош с математической точки зрения, но чувствителен к случайным отклонениям выходных параметров Y_j и приводит к смещению уравнения регрессии в их сторону. Для того чтобы минимизировать влияние случайных отклонений может быть использован критерий MAE (Mean Absolute Error). Задача с этим критерием может быть представлена следующим образом:

$$\min \sum_{j \in J} \left| \sum_{i \in I} X_{ij} * a_i + b - Y_j \right| \quad (3)$$

Комбинация моделей (1) и (3) приведет к Huber regression.

Если признаки сильно коррелированы (мультиколлинеарны), то необходимо предусмотреть в модели возможность исключения мультиколлинеарных признаков за счет правильного выбора функцией цели. Для этого существует 3 вида моделей регрессии: гребневая (Ridge), лассо (Lasso) и эластичная (Elastic net). Все эти регрессии совмещают в одной модели 2 критерия.

В случае гребневой регрессии целевая функция будет следующего вида:

$$\min \sum_{j \in J} \left(\sum_{i \in I} X_{ij} * a_i + b - Y_j \right)^2 + L * \left(\sum_{i \in I} a_i^2 + b^2 \right) \quad (4)$$

где L – большая константа (коэффициент регуляризации).

В случае регрессии лассо величины a_i и b регулируются другим способом:

$$\min \sum_{j \in J} \left(\sum_{i \in I} X_{ij} * a_i + b - Y_j \right)^2 + L * \left(\sum_{i \in I} |a_i| + |b| \right) \quad (5)$$

При объединении гребневой и лассо регрессий получаем эластичную. Так как изменением L можно регулировать величины a_i и b , данные модели часто называют моделями регу-

ляризации. Все вышеуказанные модели регрессии представлены как задачи МП и могут быть решены современными пакетами (CPLEX, Gurobi, Mosek Xpress, и т.д.), многие из которых имеют соответствующее API Python. Примеры использования МП в регрессионном анализе приводятся в ряде статей других авторов. Среди иностранных исследований интересный подход к проблеме можно выделить в работах [6 – 8]. Российские авторы также имеют значимые научные результаты в данном направлении, такие как [9, 10].

Модификации моделей ЛР, возникшие при решении практических задач

В данной статье приведена только часть примеров, с которыми авторы столкнулись на практике. Все они возникли в результате выполнения пожеланий заказчиков исследований и в дальнейшем могут возникнуть новые постановки, к которым надо быть готовым. Обычно заказчиками исследований выступают практические специалисты высокой квалификации в своей области. Их исследования направлены на решение конкретных задач и им требуется, чтобы их пожелания были выполнены или аргументировано доказаны, что они невыполнимы. Причем вне зависимости от наличия опыта использования МО в исследованиях практические специалисты сразу видят некоторые болевые точки в данных. К их числу, в частности, относятся случайные выбросы в данных. Во многих случаях отфильтровать такие наблюдения заранее не представляется возможным, поэтому модель сама должна исключать их автоматически и выдавать решения без учета влияния исключенных наблюдений. Конечно, список исключенных наблюдений предоставляется заказчику исследования для дополнительного анализа, почему такие выбросы возникли. Причины могут быть разными от элементарных ошибок при регистрации данных, сбоев оборудования до появления новых тенденций, которые могут быть предметом отдельного исследования.

Для исключения подозрительных наблюдений может быть использована следующая модель:

$$\min \sum_{j \in J} z_j \quad (6)$$

$$(1 - d) * Y_j - L * z_j \leq \sum_{i \in I} X_{ij} * a_i + b \leq (1 + d) * Y_j + L * z_j \quad j \in J \quad (7)$$

где d – задаваемая константа в диапазоне $0 < d < 1$ (максимально допустимый диапазон отклонения прогнозируемых параметров от выходных); z_j – признак непопадания или попадания в диапазон d (булева переменная).

В модели (6) – (7) минимизируется количество наблюдений, не попадающих в заданный диапазон отклонений. Конечно, можно использовать более сложные целевые функции, например:

$$\min \sum_{j \in J} z_j * L + \sum_{j \in J} \left| \sum_{i \in I} X_{ij} * a_i + b - Y_j \right| \quad (8)$$

Именно такие модели, использовались при решении задач «Учет в модели ограничения по КРІ, установленного руководством» и «Анализ и прогнозирование выхода годного и прочности агломерата» по [11, 12]. Такие модели называются MILP (Mixed Integer Linear Programs) regression. Естественно, данный подход необходимо сравнить с другими моделями регрессии. Такое сравнение приведено в [12].

В ряде случаев мы использовали следующую последовательность действий:

1. Решали задачу на основе модели (6), (7), т.е. устраняли большие отклонения.
2. Отфильтровывали случайные выбросы.
3. На уточненном множестве использовали модель (1) или (3).

Отметим, что такой подход к решению задач регрессии является удобным с практической точки зрения, так как позволяет, меняя значения допустимого диапазона, оценивать какое количество наблюдений необходимо исключить из рассмотрения. Естественным образом сразу возникает другая постановка задачи: каким может быть минимальный диапазон откло-

нений при количестве исключенных наблюдений не выше заданного уровня. Для решения такой задачи достаточно изменить целевую функцию и добавить одно ограничение к системе ограничений (7). Приведем эти дополнительные условия:

$$\min d \quad (9)$$

$$\sum_{j \in J} z_j \leq G \quad (10)$$

где G – верхняя граница количества наблюдений, исключенных из рассмотрения.

Вообще, при решении практических задач могут возникать достаточно оригинальные постановки. Например, при прогнозировании фазового состава агломерата в зависимости от изменения технологических параметров возникла необходимость использовать модель кусочно-линейной регрессии. После получения первых результатов возникла новая постановка: сколько и каких функций (не только линейных) одновременно присутствуют в выборке и можно ли найти их в рамках одной модели. Для данных целей была использована следующая модель:

$$\min \sum_{j \in J} \omega_j \quad (11)$$

$$(1 - d) * Y_j - L * z_j^f \leq \sum_{i \in I} X_{ij} * a_i^f + b^f \leq (1 + d) * Y_j + L * z_j^f \quad f \in F, j \in J \quad (12)$$

$$\sum_{f \in F} z_j^f \leq k - 1 + \omega_j \quad j \in J \quad (13)$$

где k – количество функций; F – множество функций; f – индекс функции $f = 1 \dots k$; a_i^f – коэффициенты при параметрах функций; b^f – свободные члены функций; z_j^f – признак попадания в диапазон f -ой функции (булева переменная); ω_j – признак непопадания в диапазон всех функций (булева переменная).

Минимизация числа признаков и учет дополнительных требований к коэффициентам уравнений регрессии на основе моделей MILP

При решении задач регрессии практически всегда требуется минимизировать число признаков в уравнениях регрессии. В классических моделях регрессии для этого обычно используется метод Лассо. Естественно, что для этих целей могут быть использованы и булевы переменные. Различные разновидности таких моделей с момента их возникновения достаточно подробно описаны в работах Базилевского М.П. [9].

В приведенных нами моделях достаточно часто используются булевы переменные для минимизации числа признаков и обычно это делается в 4 этапа:

1. Решаем серию задач на основе модели (6), (7) при различных значениях d и определяемся со значениями d и G .

2. Ограничиваем условием (10) предельно допустимое количество наблюдений, которые могут быть исключены из рассмотрения.

3. Вводим следующие новые дополнительные булевы переменные и ограничения на их основе:

$$-L * z_i \leq a_i \leq L * z_i \quad (14)$$

где z_i – булева переменная, при 1 информативен, при 0 неинформативен.

4. В качестве целевой функции используем:

$$\min \sum_{i \in I} z_i \quad (15)$$

То есть в окончательном виде модель минимизации числа признаков выглядит как (7), (10), (14) – (15). Данный подход был использован при прогнозировании прочностных характеристик агломерата [8].

Исходный набор данных составлял 3200 наблюдений при 240 параметрах и многократно разбивался на тренировочную выборку (2500 наблюдений) и тестовую (700 наблюдений). Для поиска уравнений регрессии в качестве базовой использовалась модель (7), (10), (14), (15). В качестве информативных параметров были оставлены 43 параметра, и практические специалисты смогли обосновать почему именно эти параметры подходят наилучшим образом, т.е. смогли их проинтерпретировать. Отметим, что интерпретируемость полученных решений было обязательным требованием заказчиков исследования, и не интерпретируемые подходы к решению задач регрессии (нейронные сети, случайный лес и т.п.) нам не подходили. Более того, по ходу решения задачи возникали дополнительные требования к коэффициентам уравнений регрессии. Данные требования и интерпретация полученных решений описаны нами в [8, 9].

Прогнозирование нескольких выходных параметров минимальным набором одинаковых входных параметров

Начнем с того как такая постановка возникла. Для оценки качества агломерата в агломерационном цехе используется 2 показателя: холодная прочность и транспортная. Холодная прочность показывает сколько мелкой фракции образуется при производстве агломерата, а транспортная при его транспортировке по конвейеру до доменной печи. Наличие мелкой фракции в доменной шихте отрицательно влияет на работу доменной печи. Поэтому количество мелкой фракции необходимо уметь прогнозировать и понимать, какие факторы наиболее влияют на ее образование. Возникает вопрос: «Зачем прогнозировать 2 показателя одним набором признаков?». Гораздо проще найти уравнение регрессии для каждого показателя по отдельности. Кроме того, наборы информативных признаков для различных показателей могут отличаться друг от друга. Согласны, но для сбора данных используются определенные датчики, причем некоторые из них достаточно дорогостоящие. Поэтому необходимо выяснить, существует ли набор датчиков, который с заданной точностью позволит прогнозировать 2 различных показателя.

Для построения модели, в рамках которой будем прогнозировать несколько выходных параметров с минимальным набором одинаковых входных параметров, будем использовать следующую систему ограничений:

$$(1 - d) * Y_j^t - L * z_j^t \leq \sum_{i \in I} X_{ij} * a_i^t + b^t \leq (1 + d) * Y_j^t + L * z_j^t \quad t \in T, j \in J \quad (16)$$

$$\sum_{j \in J} z_j^t \leq G^t \quad t \in T \quad (17)$$

$$-L * z_i \leq a_i^t \leq L * z_i \quad t \in T \quad (18)$$

$$\min \sum_{i \in I} z_i \quad (19)$$

Множество T в данной модели – это множество выходных параметров. Смысл остальных обозначений тот же что и в предыдущих моделях. В модели (16) – (19) допускается, что при поиске разных уравнений регрессии могут быть исключены разные наборы наблюдений. Для того, чтобы исключался один и тот же набор наблюдений, необходимо ужесточить ограничения (16), (17). При излишнем ужесточении системы ограничений может возникать противоречивость. В таком случае необходимо ослабить некоторые ограничения введением переменных их корректирующих и минимизацией числа таких корректировок.

Заключение

В данной статье приведен ряд моделей регрессии, которые использовались на практике и дано описание как возникали те или иные ситуации. Естественно, нельзя предугадать другие ситуации и требования заказчиков исследования, которые могут возникнуть. Поэтому приведенные модели лучше воспринимать как некоторый конструктор, из которого всегда можно сделать необходимую модель.

Еще хотелось бы отметить один очевидный факт, что готовые уравнения ЛР достаточно легко можно использовать в различных моделях, относящихся к исследованию операций (ИО), особенно если это тоже модели МП. Более того, во многих случаях уравнения ЛР именно для этого и предназначены. Более интересным является то, что модель (16) – (19) в исходном виде может быть объединена с моделями ИО. Обратите внимание на текст, выделенный в данном абзаце жирным шрифтом. Если выбор того или иного подмножества признаков в качестве наиболее информативного влечет за собой определенный набор организационных мероприятий, то их можно оценить в стоимостном выражении, а это означает, что модель можно развивать дальше за счет ввода дополнительных ограничений, не имеющих никакого отношения к МО, или видоизменить функцию цели (19). В этом случае в ней перед z_i появятся стоимостные коэффициенты, и в модели можно будет учитывать экономические показатели.

Предлагаемый подход обладает следующими достоинствами:

1. Позволяет формировать модели регрессии с учетом требований, сформулированных заказчиком. Если достижение целевых установок заказчика невозможно, то предусмотрена возможность их оптимальная корректировки в рамках модели.

2. Позволяет минимизировать число признаков необходимых для выполнения первого пункта.

3. Математические модели достаточно простые, хорошо воспринимаются практическими специалистами, не требуют знания методов МП и навыков программирования на языках высокого уровня.

4. Модели ЛР на основе МП легко встраиваются в другие модели ИО.

Критики нашего подхода обычно утверждают, что задачи MILP плохо и долго считаются. После этого упоминаются NP-полные задачи и т.п. Конечно, алгоритм с полиномиальным изменением времени счета в зависимости от размерности задачи для решения NP-полных задач до сих пор не найден, но из этого не следует, что не надо решать задачи MILP. С момента появления пакета MPSX-360, а это начало 80-х годов прошлого века, задачи MILP большой размерности успешно решаются. Действительно, время счета задач с сотнями переменных и десятками тысяч наблюдений может составлять несколько часов, но на выходе получается качественное решение, соответствующее требованиям заказчика исследования.

Список источников:

1. Гельман Э., Хилл Д., Вехтари А. Регрессия: теория и практика. С примерами на R и Stan // М.: ДМК пресс, 2022. – 748 с.
2. Ye A. Machine Learning Regression Algorithms You Need to Know, 2020, URL: <https://medium.com/analytics-vidhya/5-regression-algorithms-you-need-to-know-theory-implementation-37993382122d> (дата обращения 08.12.20220)
3. Willmott C.J., Matsuura Kenji Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Research*, 2005, 30 p. 79-82.
4. Huber, P.J (1964) Robust Estimation of a Location Parameter. *Annals of Statistics*, 53(1), p. 73-101.
5. Zou, H. & Hastie, T. Regularization and Variable Selection via the Elastic Net. *Journal of the Royal Statistical Society Series B*, 2005, p. 301-320
6. Gkioulekas Y. & Papageorgiou L.G. Piecewise regression through the Akaike information criterion using mathematical programming. *IFAC PapersOnLine*, 2018, vol. 51, no. 15, pp. 730-735.
7. Wang B.Q., Chukova S., Lai C.D. On the relationship between regression analysis and mathematical programming. *Journal of applied mathematics and decision sciences*, 2004, vol. 8, no. 2, pp. 131-140.

References:

1. Gelman E., Hill D., Vekhtari A. Regression: Theory and Practice. With Examples in R and Stan. Moscow: DMK Press; 2022.
2. Ye A. Machine Learning Regression Algorithms You Need to Know [Internet]. 2020 [cited 2022 Dec 08]. Available from: <https://medium.com/analytics-vidhya/5-regression-algorithms-you-need-to-know-theory-implementation-37993382122d>
3. Willmott C.J., Matsuura K. Advantages of the Mean Absolute Error (MAE) Over the Root Mean Square Error (RMSE) in Assessing Average Model Performance. *Climate Research*. 2005;30:79-82.
4. Huber P.J. Robust Estimation of a Location Parameter. *Annals of Statistics*. 1964;53(1):73-101.
5. Zou H., Hastie T. Regularization and Variable Selection via the Elastic Net. *Journal of the Royal Statistical Society. Series B*. 2005:301-320.
6. Gkioulekas Y., Papageorgiou L.G. Piecewise Regression Through the Akaike Information Criterion Using Mathematical Programming. *IFAC PapersOnLine*. 2018;51(15):730-735.
7. Wang B.Q., Chukova S., Lai C.D. On the Relationship Between Regression Analysis and Mathematical Programming. *Journal of Applied Mathematics and Decision Sciences*. 2004;8(2):131-140.

8. Yang, L., Liu, S., Tsoka, S. & Papageorgiou, L.G. A regression tree approach using mathematical programming. *Expert Systems with Applications*, 2017, vol. 78, pp. 347-357.

9. Базилевский М.П. Сведение задачи отбора информативных регрессоров при оценивании линейной регрессионной модели по методу наименьших квадратов к задаче частично-булевого линейного программирования // *Моделирование, оптимизация и информационные технологии (Воронеж)*, 2018. – Т.6. – №1 (20). – С. 108-117.

10. Горелика В.А., Трембачева О.С. Решение задачи линейной регрессии с использованием методов матричной коррекции в метрике // *Ж. вычисл. матем. и матем. физ.*, 2016. – Т. 56. – № 2. – С. 202-207.

11. Чернавин П. Ф., Малыгин А. В., Деткова Т. В., Кучин В.Ю. Анализ и прогнозирование выхода годного и прочности агломерата на основе моделей математического программирования // *Черные Металлы*, 2021. – № 12. – С. 20-24.

12. Чернавин П.Ф., Гайнанов Д. Н., Панкрасченко В.Н., Чернавин Ф.П., Чернавин Н.П. Машинное обучение на основе задач математического программирования // *М.: Наука*, 2021. – 128 с.

Информация об авторах:

Чернавин Павел Федорович – кандидат экономических наук, +79120414084, доцент кафедры Аналитика больших данных и методы видеоанализа, AuthorID: 117430, SPIN:6370-8103, ORCID: 0000-0003-3214-3906

Чернавин Николай Павлович - ассистент кафедры Аналитика больших данных и методы видеоанализа, AuthorID: 971565, SPIN:5722-9436, ORCID: 0000-0003-3214-3906

Чернавин Федор Павлович – кандидат экономических наук, +79120504811, доцент кафедры Моделирование управляемых систем, AuthorID: 971595, SPIN: 9237-5190, ORCID: 0000-0003-3214-3906

8. Yang L., Liu, S., Tsoka S., Papageorgiou L.G. A Regression Tree Approach Using Mathematical Programming. *Expert Systems With Applications*. 2017;78:347-357.

9. Bazilevsky M.P. Reduction the Problem of Selecting Informative Regressors When Estimating a Linear Regression Model by the Method of Least Squares to the Problem of Partial-Boolean Linear Programming. *Modeling, Optimization and Information Technologies (Voronezh)*. 2018;6-1(20):108-117.

10. Gorelika V.A., Trembacheva O.S. Solution of a Linear Regression Problem Using Matrix Correction Methods in Metrics. *Computational Mathematics and Mathematical Physics*. 2016;56(2):202-207.

11. Chernavin P.F., Malygin A.V., Detkova T.V., Kuchin V.Yu. Analysis and Prediction of Sinter Yield and Strength Based on Mathematical Programming Models. *Chernye Metally*. 2021;12:20-24.

12. Chernavin P.F., Gainanov D.N., Pankraschenko V.N., Chernavin F.P., Chernavin N.P. Machine Learning Based on Mathematical Programming Problems. *Moscow: Nauka*; 2021.

Information about authors:

Chernavin Pavel Fedorovich – Candidate of Economic Sciences, Associate Professor of the Department «Big Data Analytics and Video Analysis Methods», AuthorID: 117430, SPIN: 6370-8103, ORCID: 0000-0003-3214-3906

Chernavin Nikolai Pavlovich – Assistant of the Department «Big Data Analytics and Video Analysis Methods», AuthorID: 971565, SPIN: 5722-9436, ORCID: 0000-0003-3214-3906

Chernavin Fedor Pavlovich – Candidate of Economic Sciences, ph. +79120504811, Associate Professor of the Department «Controlled Systems Modelling», AuthorID: 971595, SPIN: 9237-5190, ORCID: 0000-0003-3214-3906

Вклад авторов: все авторы сделали эквивалентный вклад в подготовку публикации.

Contribution of the authors: the authors contributed equally to this article.

Авторы заявляют об отсутствии конфликта интересов.

The authors declare no conflicts of interests.

Статья поступила в редакцию 09.01.2023; одобрена после рецензирования 08.02.2022; принята к публикации 26.02.2023.

The article was submitted 09.01.2023; approved after reviewing 08.02.2022; accepted for publication 26.02.2023.

Рецензент – Подвесовский А.Г., кандидат технических наук, доцент, Брянский государственный технический университет.

Reviewer – Podvesovskij A.G., Candidate of Technical Sciences, Associate Professor, Bryansk State Technical University.