

Математическое моделирование

УДК 004.855.5

DOI:10.25729/ESI.2023.30.2.005

Оптимизационные модели подбора параметров технологических процессов на основе результатов машинного обучения

Чернавин Павел Федорович, Чернавин Николай Павлович, Чернавин Федор Павлович
Уральский федеральный университет, Россия, Екатеринбург, *p.f.chernavin@urfu.ru*

Аннотация. При решении практических задач достаточно часто возникает необходимость одновременного применения методов машинного обучения и методов исследования операций. Так как в основе многих методов решения задач в обеих областях может лежать кардинально отличающийся друг от друга математический инструментарий, то объединить их результаты в единый комплекс будет невозможно. В данной статье приведен взаимосвязанный комплекс моделей машинного обучения и исследования операций, предназначенный для подбора параметров технологических процессов. Все модели имеют общую математическую базу на основе задач математического программирования с частично-целочисленными переменными. Модели апробированы на реальной задаче подбора состава шихты и технологических параметров агломерационного производства и приведены в последовательности их возникновения в процессе решения задач, поставленных заказчиками исследования. На первом этапе на основе задач регрессии осуществляется отбор наиболее информативных входных признаков и степень их влияния на выходные признаки. Затем на основе задач классификации определяются рекомендуемые области управляемых входных признаков для получения высококачественной продукции. Данные области могут иметь достаточно сложную геометрическую конфигурацию в пространстве признаков. Далее в рамках задач исследования операций определяются эталонные состояния процесса, к которым необходимо стремиться. На заключительном этапе результаты всех предыдущих исследований объединяются в единую оптимизационную модель, которая может быть дополнена результатами исследований, полученных из других источников информации, если эти результаты можно представить как линейные ограничения. Предлагаемый подход к оптимизации параметров может быть использован в различных предметных областях.

Ключевые слова: математическое программирование, машинное обучение, исследование операций, регрессия, классификация, прочность агломерата

Цитирование: Чернавин П.Ф. Оптимизационные модели подбора параметров технологических процессов на основе результатов машинного обучения / П.Ф. Чернавин, Н.П. Чернавин, Ф.П. Чернавин // Информационные и математические технологии в науке и управлении. – 2023. – № 2(30). – С. 45-56 – DOI:10.25729/ESI.2023.30.2.005.

Введение. Существуют различные направления математического моделирования. В данной статье нами рассмотрены только два из них: исследование операций (ИО) и машинное обучение (МО). На наш взгляд, в настоящий момент в математическом моделировании технологических процессов существует некоторый разрыв между использованием моделей машинного обучения (МО) и моделей исследования операций (ИО). С одной стороны, любая из этих моделей имеет собственную ценность, и они могут применяться в различного рода исследованиях независимо друг от друга. С другой стороны, например, уравнения регрессии предназначены для прогнозирования выходного признака некоторого процесса на основании входных признаков. Возникает простой вопрос: прогноз выходного признака нужен сам по себе или для того, чтобы понять, в какую сторону будет двигаться технологический процесс при изменении входных признаков. Аналогичная ситуация и с задачами классификации. Для простоты, пусть есть 2 класса: хорошее состояние процесса и неудовлетворительное. Конечно, важно уметь по входным признакам понимать состояние процесса, но, наверное, более важно понять, что надо делать, если процесс движется в неблагоприятную сторону. Как говорят философы: «Мало понять почему, главное, как изменить в лучшую сторону». Хотя и понять, почему, тоже не просто.

Ответ на вопрос: «Почему?», – могут дать хорошо интерпретируемые модели МО, на вопрос: «Как изменить что-то в лучшую сторону?», – модели ИО. Таким образом, ответ на оба вопроса в большинстве случаев можно получить, только объединив два этих подхода. К сожалению, многие модели МО (нейронные сети, градиентный бустинг, случайный лес и т.п.) либо слабо интерпретируемые, либо не интерпретируемые вообще. Скажем больше, даже ряд хорошо интерпретируемых моделей (методы ближайших соседей, потенциальных функций, деревьев решений) практически невозможно совместить с моделями ИО.

На наш взгляд, одним из условий, необходимых для возможности совмещения моделей ИО и МО, является наличие общей математической базы и способов представления результатов решений, и сделать это можно на основе моделей математического программирования (МП). Действительно, многие методы МО, такие, как классическая линейная регрессия и ее различные модификации (лассо, гребневая, эластичная), линейное разделение множеств и ансамбли на базе линейных разделителей, метод опорных векторов, могут быть представлены, как задачи МП [1-10]. Причем решения, полученные на их основе, являются обычными линейными уравнениями и поэтому легко могут быть включены в различные модели ИО. Чтобы не быть голословными, поделимся нашим опытом интеграции моделей МО и ИО в единый комплекс при решении конкретной практической задачи управления технологическими параметрами для получения высококачественного агломерата. Данная задача решалась по заказу ПАО «Северсталь» на основе реальных данных агломерационного цеха №3.

1. Описание технологического процесса. Производство агломерата стоит в самом начале металлургического процесса. Поэтому от его технико-экономических показателей зависит эффективность всех последующих переделов. Агломерация – сложный, многомерный, растянутый во времени технологический процесс. На качественные характеристики агломерата влияет большое количество различных факторов: химико-минералогический и гранулометрический состав сырья, условия дозирования, увлажнения, смешивания и укладки шихты, режимы спекания и охлаждения агломерата, сбой оборудования и ошибки персонала.

Основной проблемой при производстве агломерата является выход мелкой фракции. Ее нельзя использовать в доменном производстве и поэтому ее приходится перерабатывать вновь. На основе количества мелкой фракции в агломерационном производстве рассчитывается показатель «холодная прочность» агломерата. На основе количества мелкой фракции, образующейся при транспортировке до доменного цеха, рассчитывается показатель «транспортная прочность» агломерата. Высококачественный агломерат должен одновременно соответствовать как минимум этим двум показателям. Для исследований нам был предоставлен набор из 3200 наблюдений. Каждое наблюдение содержало 321 входной признак. Далее опишем, в какой последовательности возникали взаимосвязанные по смыслу, но разные по методам решения задачи, и как они были решены.

2. Первоначальная постановка задачи. В начале исследований была поставлена задача – найти уравнения регрессии, прогнозирующие холодную и транспортную прочность при минимальном числе необходимых для этого признаков. Стандартными методами машинного обучения на основе программ из библиотеки Sklearn была сделана проверка на мультиколлинеарность и информативность признаков. В результате количество признаков было уменьшено до 240. После обсуждения с практическими специалистами информативности признаков были оставлены 78 признаков и именно на их основе строились модели регрессии.

Как уже отмечалось, агломерация относится к числу сложных и многомерных процессов, поэтому на входные и выходные признаки влияют множество факторов. По этой причине в исходных данных могут содержаться нетипичные наблюдения (выбросы), устранение которых простыми статистическими методами невозможно. Термин «выбросы» (outliers) является устоявшимся термином, поэтому далее по тексту мы будем его использовать. Изначальным

требованием к модели регрессии было автоматическое исключение выбросов в рамках самой модели. Для этих целей была использована следующая модель:

$$\min \sum_{j \in J} z_j, \quad (1)$$

$$(1 - d) * Y_j - L * z_j \leq \sum_{i \in I} X_{ij} * a_i + b \leq (1 + d) * Y_j + L * z_j, \quad j \in J, \quad (2)$$

где I – множество исходных входных признаков;

J – множество наблюдений;

X_{ij} – значения входных признаков;

Y_j – значения выходного признака;

a_i – коэффициенты уравнения регрессии (искомые переменные);

b – свободный член уравнения регрессии для t -ого показателя (искомая переменная);

d – число ($0 < d < 1$);

z_j – булева переменная;

L – большая константа, значительно превышающая входные и выходные параметры.

Работа с моделью (1) – (2) осуществляется следующим образом. Задается величина d в рамках указанного диапазона. Булевы переменные предназначены для фиксации факта попадания или нет конкретного наблюдения в диапазон, задаваемый числом d . Наблюдения, у которых $z_j = 1$, расцениваются, как выбросы. Таким образом, модель (1) – (2) позволяет найти разумный компромисс между погрешностью прогноза выходных признаков и количеством наблюдений, исключенных из рассмотрения. Конечно, список исключенных из рассмотрения наблюдений анализировался отдельно на предмет обоснованности их исключения. В кодах пакета IBM ILOG CPLEX описание данной модели приведено на листинге 1:

Листинг 1. Модель регрессии с исключением выбросов в рамках заданного диапазона.

```
float d=...; //число, задающее диапазон погрешности прогноза (доля от
выходного признака)
int L=...; //большое число
int k=...; //число наблюдений
range j = 1..k; // индекс наблюдения
int n=...; //число входных признаков
range i = 1..n; // индекс признака
float X[j][i]=...; //множество значений входных признаков
float Y[j]=...; //множество значений выходного признака
/* искомые переменные */
dvar float a[i]; //коэффициенты уравнения регрессии
dvar float b; //свободный член уравнения регрессии
dvar boolean z[j]; //признак попадания в диапазон
/*целевая функция*/
Minimize sum(j in j) z[j];
/*система ограничений*/
subject to {
forall( j in j) sum(i in i) X[j][i]*a[i]+b <= (1+d)*Y[j]+L*z[j];
forall( j in j) sum(i in i) X[j][i]*a[i]+b >= (1-d)*Y[j]-L*z[j];
};
```

На основе модели (1) – (2) были сформулированы уравнения регрессии для прогноза четырех выходных параметров. Задачи решались для каждого выходного параметра отдельно. Значение d задавалось в диапазоне от 0.05 до 0.15 с шагом 0.01. Требовалось найти разумный компромисс между отклонением прогнозируемых показателей от фактических и процентом исключенных из рассмотрения наблюдений. Результаты расчетов оценивались практическими специалистами по агломерационному производству. По их мнению, разумный компромисс достигается при $d=0.1$, при этом процент исключенных из рассмотрения наблюдений составил

от 5.47% до 11.04%. Отметим, что список исключенных наблюдений специалистами анализировался отдельно для выяснения причин отклонений. Более подробно описание расчетов приведено в [11]. Там же приведено сравнение качественных характеристик предлагаемого подхода с другими моделями регрессии. Расчеты проводились на компьютерах с частотой процессора 4.00 GHz и объемом оперативной памяти 32.0 ГБ. Время счета отдельной задачи составляло 1.5-3.5 часа и зависело от выходного параметра и величины d .

В рамках модели (1)-(2) мы нашли устраивающие нас значения d и $\sum_{j \in J} z_j$. Обозначим их как d^* и G . Тогда, чтобы избавиться от неинформативных признаков, нам надо решить следующую задачу:

$$\min \sum_{i \in I} w_i, \quad (3)$$

$$(1 - d^*) * Y_j - L * z_j \leq \sum_{i \in I} X_{ij} * a_i + b \leq (1 + d^*) * Y_j + L * z_j, \quad j \in J, \quad (4)$$

$$\sum_{j \in J} z_j \leq G, \quad (5)$$

$$-L * w_i \leq a_i \leq L * w_i, \quad i \in I. \quad (6)$$

где w_i - булева переменная (1 – признак информативен, 0 – неинформативен).

Естественно, что программу в кодах CPLEX необходимо дополнить соответствующими условиями и изменить целевую функцию. Если расчеты по моделям (1) – (2) и (3) – (6) делать отдельно для каждого выходного признака (в нашем случае их было два), то может возникнуть ситуация, когда наборы информативных признаков для различных выходных признаков не будут совпадать. Если требуется подобрать минимальный набор информативных входных признаков, который подходит для прогнозирования нескольких выходных признаков, то условия (4) и (6) следует заменить на следующие:

$$(1 - d^t) * Y_j^t - L * z_j \leq \sum_{i \in I} X_{ij} * a_i^t + b^t \leq (1 + d^t) * Y_j^t + L * z_j, \quad t \in T, j \in J, \quad (4.1)$$

$$-L * w_i \leq a_i^t \leq L * w_i, \quad i \in I, t \in T. \quad (6.1)$$

где T – множество выходных показателей (холодная и транспортная прочности);

Y_j^t – t -ый выходной признак j -го наблюдения (числа);

a_i^t – коэффициенты уравнения регрессии для t -ого показателя (искомые переменные);

b^t – свободный член уравнения регрессии для t -ого показателя (искомая переменная);

d^t – числа ($0 < d^t < 1$).

Результаты расчетов по модели (3), (4.1), (5), (6.1) и таблица сравнения их качества с другими моделями регрессии приведены в [11]. Там же приведен список наиболее информативных параметров и степень их влияния на различные показатели качества агломерата. Из 78 исходных признаков на основе моделей (3) – (6) в качестве информативных были выбраны 43 признака. Причем практическим специалистам агломерационного цеха представлялись наборы признаков одинаковые по их числу, но отличающиеся по составу. Делалось это с целью отобрать наиболее интерпретируемые и управляемые, чтобы в дальнейшем можно было влиять на процесс агломерации и понимать к чему должно привести изменение тех или иных параметров. В данной статье подробно не рассматривается вопрос интерпретируемости признаков, но хотелось бы отметить, что интерпретируемость решений стала настолько важной, что в исследованиях по искусственному интеллекту (ИИ) уже формируется новое направление: интерпретируемый ИИ.

3. Определение областей управляемых параметров для получения высококачественного агломерата. Практическим специалистам модели регрессии были необходимы для

оценки влияния входных параметров на величину выходных, но на основе уравнений регрессии достаточно сложно определить допустимые области входных параметров, нахождение в которых обеспечивает получение высококачественного агломерата. Причем допустимые области могут иметь достаточно сложную конфигурацию, которую невозможно описать только через нижние и верхние границы по каждому параметру в отдельности.

Для начала требовалось определиться, что будем понимать под высококачественным агломератом и агломератом удовлетворительного качества. Практическими специалистами были определены верхние границы по выходу мелочи в агломерационном цехе и при транспортировке агломерата в доменный цех. Если агломерат одновременно удовлетворял двум этим показателям, то он считался высококачественным, по этому принципу наблюдения были разбиты на 2 класса. Высококачественный класс составил 535 наблюдений, удовлетворительный – 2680 наблюдений. Отметим, что заказчиков исследования интересовало не просто решающее правило (РП), которое давало бы ответы «да», «нет», а именно области в пространстве признаков, соответствующие этим классам. Кроме того, при выборе метода поиска РП следовало учесть, что классы сильно не сбалансированы. Для решения поставленной задачи была использована авторская математическая модель комитетных конструкций в наиболее общем виде. Комитетный подход для решения задач классификации наиболее активно развивался в Уральской школе распознавания образов под руководством Мазурова В. Д. [9, 10]. Важнейшим качеством метода комитетов является то обстоятельство, что полученные результаты имеют геометрическую интерпретацию, а РП может быть представлено в виде системы линейных ограничений, в которую в дальнейшем можно включать дополнительные условия. Комитетный подход к решению задач классификации на основе моделей линейного программирования с частично целочисленными переменными достаточно подробно описан авторами статьи в монографии [12]. В данной статье приводится модель, где условия разграничения множеств гиперплоскостями были записаны следующим образом:

$$\sum_{i \in I} X_{ij} * a_i^h + b^h + L * z_j^h \geq E, \quad j \in J_1, \quad h \in H, \quad (7)$$

$$\sum_{i \in I} X_{ij} * a_i^h + b^h - L * z_j^h \leq 0, \quad j \in J_2, \quad h \in H, \quad (8)$$

где J_1 и J_2 – множества, которые необходимо разделить;

$I = J_1 \cup J_2$ – множество наблюдений;

H – множество членов комитета (гиперплоскостей, линейных нейронов);

h – индекс гиперплоскости;

$a_i^h x_i^t$ – коэффициент гиперплоскости h для i -го параметра (переменная);

b^t – свободные члены гиперплоскостей (переменная);

$z_j^h z_j^t$ – булевы переменные для фиксации нарушений условий разделения множеств;

E – малое число (используем для строгости ограничений).

В контексте данной статьи термины «разделяющая гиперплоскость», «член комитета» и «линейный нейрон» используются как синонимы. В модели (7), (8) одновременно присутствует L – большое число и E – малое число. В ряде программ для решения задач МП, которые использовались в прошлом веке, это могло породить проблему потери точности счета. В современных пакетах программ такая проблема практически отсутствует.

В практических задачах обычно не требуется построение комитета со 100 % разделяющей способностью. Более того, излишне высокая точность решающего правила (РП) может свидетельствовать о переобученности РП. Поэтому условия комитета сразу записываем с возможностью их корректировки:

$$\sum_{h \in H} V^h * z_j^h \leq m + \left(\sum_{h \in H} V^h \right) * \Delta_j, \quad j \in J_1, \quad (9)$$

$$\sum_{h \in H} V^h * z_j^h \leq \sum_{h \in H} V^h - m - 1 + \left(\sum_{h \in H} V^h \right) * \Delta_j, \quad j \in J_2, \quad (10)$$

$$\text{Целевая функция: } \min \sum_{j \in J} \Delta_j, \quad (11)$$

где V^h – веса членов комитета (константы);

m – квалифицированное меньшинство (переменная $0 \leq m \leq \sum_{h \in H} V^h - 1$);

Δ_j – булевы переменные для фиксации нарушений условий комитета.

В качестве весов удобно использовать степени числа 2, то есть ($V^h = 2^{h-1}$; $h = 1, 2, \dots, H^*$, где H^* – задаваемое число гиперплоскостей). Модель (7) – (11) является универсальной и на ее основе могут быть построены комитеты единогласия, большинства и старшинства. Подробное описание каждого из этих комитетов с графическими иллюстрациями приведено в [12]. В кодах CPLEX описание данной модели приведено на листинге 2:

Листинг 2. Универсальная модель комитетов.

```

/*описание данных*/
int N=...; //число гиперплоскостей (линейных нейронов, членов комитета)
range h = 1..N; //индекс члена комитета
int V[h]=...; // веса членов комитета
int SV=...; //сумма весов (гиперплоскостей, членов комитета)
int N=...; //число входных признаков
range i = 1..N; // индекс входного признака
int K1=...; //число наблюдений в множестве 1
int K2=...; //число наблюдений в множестве 2
range j1 = 1..K1; // индекс наблюдения 1-го множества
range j2 = 1..K2; // индекс наблюдения 2-го множества
int L=...; //большое число для коррекции условий разделения множеств
// в случае невозможности выполнения этих условий
float E=...; //малое число для выполнения строгости ограничений
float X1[j1][i]=...; //значение признаков наблюдений множества 1
float X2[j2][i]=...; //значение признаков наблюдений множества 2
/*искомые переменные*/
dvar float a[i][h]; //коэффициенты гиперплоскостей
dvar float b[h]; //свободные члены гиперплоскостей
dvar boolean z1[j1][h]; // индикатор нарушения условия отнесения к множеству1
dvar boolean z2[j2][h]; // индикатор нарушения условия отнесения к множеству2
dvar boolean d1[j1]; //индикатор нарушения условий комитета для наблюдений множе-
ства1
dvar boolean d2[j2]; //индикатор нарушения условий комитета для наблюдений множе-
ства2
dvar int+ m; //меньшинство с учетом весов
/*технические переменные для нормировки коэффициентов гиперплоскостей*/
dvar boolean g[h];
dvar boolean q[h];
/*условия модели*/
minimize sum( j1 in j1)d1[j1]+sum( j2 in j2)d2[j2];
subject to {
forall (h in h)forall(j1 in j1) sum(i in i)X1[j1][i]*a[i][h]+b[h]+L*z1[j1][h] >= E;
forall(h in h)forall(j2 in j2) sum(i in i)X2[j2][i]*a[i][h]+b[h]-L*z2[j2][h]<=0;
forall(j1 in j1)sum(h in h) V[h]*z1[j1][h]<=m + L*d1[j1]; // условия комитета
forall(j2 in j2)sum(h in h)V[h]*z2[j2][h]<=SV-m-1+L*d2[j2]; // условия комитета
/*Условия для нормировки коэффициентов гиперплоскостей относительно коэффициента
при одном из признаков. Данная часть программы не является обязательной и может
быть исключена, но она очень удобна при решении практических задач*/
/*По умолчанию коэффициент при первом признаке будет = 1 или -1*/
/* Для изменения нормировки можно изменить число в [], например указать a[2][h]*/
forall(h in h) a[1][h]==g[h]-q[h];
forall(h in h) g[h]+q[h]==1;
};

```

Хорошие результаты показал комитет старшинства из 5 членов. На его основе были выделены 2 выпуклых множества в пространстве признаков, соответствующие высококачественному агломерату. В одном из них было сконцентрировано 338 наблюдений, в другом 35, остальные 162 наблюдения распределились случайным образом в пространстве признаков. Для конкретных рекомендаций было выбрано множество с наибольшим числом наблюдений. На его основе специалистами по агломерационному производству были разработаны конкретные рекомендации, которые приведены в [13].

4. Поиск эталонных точек. Выпуклые области, полученные в результате расчетов по модели (7) – (11), могут иметь достаточно сложную форму. Для лучшего восприятия результатов решения внутри этих областей могут выделяться геометрически правильные фигуры. Например, Ащепков Л.Т. в заданной области выделяет максимальный куб [14].

Так как внутри рекомендуемой выпуклой области находятся наблюдения, соответствующие высококачественному агломерату, а за ее пределами, в основном, – удовлетворительному, любой практик прекрасно понимает, что чем ближе к границам области, тем выше вероятность получить не высококачественный агломерат, а только удовлетворительного качества. Поэтому в нашем случае после того, как выпуклая область для рекомендаций была сформулирована в виде системы неравенств, у заказчиков исследования сразу возникло желание найти центр этой области. Таким центром является некоторая эталонная точка, минимальное расстояние от которой до границ области максимально. Такую точку можно найти на основе следующей модели:

$$\sum_{i \in I} a_i^h * p_i + b^h = \sqrt{\sum_{i \in I} (a_i^h)^2 * r^h}, \quad h \in H. \quad (12)$$

$$r^h \geq \varphi, \quad h \in H. \quad (13)$$

$$\max \varphi, \quad (14)$$

где a_i^h, b^h – константы, найденные на основе модели (7) – (11);

p_i – искомые координаты (технологические параметры) точки эталона;

r^h – длина перпендикуляра от эталонной точки до h -ой гиперплоскости;

φ – минимальная длина перпендикуляра до гиперплоскостей.

С геометрической точки зрения φ можно интерпретировать как радиус гипершара, вписанного в заданную выпуклую область, а координаты p_i – как координаты его центра.

В кодах CPLEX описание данной модели приведено на листинге 3:

Листинг 3. Поиск центра системы неравенств (эталонной точки).

```
int N=...; //размерность пространства (число параметров наблюдений)
range i = 1..N; // индекс параметра
int H=...; //число гиперплоскостей
range h = 1..H; // индекс гиперплоскости
float A[h][i]=...; //матрица коэффициентов гиперплоскостей
float b[h]=...; //свободные члены гиперплоскостей
float Sqrt[h]=...; //корень из суммы квадратов коэф. гиперплоскости
/* искомые переменные*/
dvar float p[i] ; //координаты центра системы (эталон)
dvar float+ r[h] ; //перпендикуляр от центра до гиперплоскости
dvar float+ minr ; //минимальное расстояние от центра до гиперплоскостей

maximize minr;
subject to {
forall(h in h) sum(i in i) A[h][i]*p[i]+b[h]==Sqrt[h]*r[h];
forall(h in h) r[h]>=minr;
};
```

Обозначим параметры эталонного состояния как $p_i^* \forall i \in I$. Точка с координатами $p_i^* \forall i \in I$ может быть рекомендована как некоторый набор параметров, к которому надо стремиться для того, чтобы устойчиво иметь хорошие качественные показатели агломерата.

Есть другой способ поиска эталонной точки. Для каждой точки, находящейся внутри рекомендуемой области, рассчитать расстояние (длина перпендикуляра) до каждой гиперплоскости и выбрать реальное наблюдение, соответствующее минимуму расстояния. Теоретикам больше нравится первый способ, практикам – второй. Естественно, что вряд ли существует некоторая универсальная точка на все случаи жизни. Правильным будет подход, когда выделяются некоторые периоды, близкие по составу шихты, сезонным и погодным условиям. Для каждого такого периода ищется эталонная точка.

5. Построение оптимизационных моделей. Современная конкурентная среда на рынке металлов постоянно побуждает к экспериментам как с составом шихты, так и с технологическими параметрами. Поэтому, после построения имитационной модели возникло желание не только автоматизировать подбор состава шихты и технологических параметров, но и расширить число учитываемых ограничений. На наш взгляд, для этих целей наилучшим образом подходят модели на основе задач МП.

Рассмотрим, какие блоки ограничений возможны для построения оптимизационной модели. В результате расчетов по модели (1) – (6) имеются коэффициенты уравнений регрессии a_i^t для каждого качественного показателя. В результате расчетов по модели (7) – (11) получена система ограничений, которая высекает допустимую выпуклую область в пространстве признаков. В результате решения задачи (12) – (14) у нас есть точка эталонного состояния.

Конечно, эталонное состояние надо знать и к нему надо стремиться, но может оказаться, что в конкретный момент времени выполнить все условия эталонного состояния невозможно. Например, в конкретный момент нет в наличии необходимого количества одного из железорудных концентратов, но есть другие, поэтому его содержание в шихте может быть в рекомендуемых границах, но меньше эталонного. По этим причинам из практических соображений следует задать границы возможного отклонения от эталонного состояния. Естественно, что эталонная точка должна находиться внутри этих границ. Тогда базовая оптимизационная модель может быть записана следующим образом:

$$\min \sum_{i \in I} (p_i^* - p_i)^2, \quad (15)$$

$$\sum_{i \in I} a_i^h * p_i + b^h \geq 0, h \in H, \quad (16)$$

$$\alpha^t \leq \sum_{i \in I} a_i^t * p_i \leq \beta^t, t \in T, \quad (17)$$

$$\alpha_i \leq p_i \leq \beta_i, i \in I, \quad (18)$$

где α^t, β^t – желаемые границы качественных показателей;

α_i, β_i – уточненные границы технологических параметров

В результате решения задачи (15) – (18) можно определить состав шихты и набор технологических параметров, соответствующих системе ограничений (16) – (18) при минимальном евклидовом расстоянии от эталонного состояния. На наш взгляд, на основе приведенных предыдущих программ в кодах CPLEX читатели смогут самостоятельно запрограммировать модель (15) – (18). Задачи квадратичного программирования пакетом IBM ILOG CPLEX решаются быстро и качественно.

Модель (15) – (18) может дополняться дополнительными ограничениями при условии, что система ограничений в целом остается выпуклым множеством. Конечно, наращивание системы ограничений в какой-то момент может привести к ее противоречивости и отсутствию

допустимого решения. В этом случае модель надо будет видоизменить и использовать последовательную оптимизацию.

Особо хотелось бы отметить, что приведенные в данной статье модели только на первый взгляд могут показаться сложными. Специалисты с инженерным образованием достаточно быстро осваивают этот инструментарий и начинают его воспринимать, как некоторый метаязык для формулировки своих пожеланий, причем специальных знаний из области математического программирования не требуется. Эффективные алгоритмы решения таких задач создавались лучшими математиками мира, начиная с тридцатых годов прошлого столетия. В настоящий момент эти алгоритмы оформлены в виде различных программных комплексов [15], которые позволяют решать задачи большой размерности за приемлемое время. Эти программные комплексы имеют дружественный интерфейс, освоить который гораздо проще, чем языки программирования высокого уровня. Для освоения формы записи требуются только логика и знание простых приемов математической записи линейных ограничений.

Приведенный в данной статье комплекс моделей был успешно использован для подбора технологических параметров в агломерационном цехе №3 ПАО «Северсталь» [11, 13]. Отметим также, что приведенные нами модели можно использовать не только для подбора параметров технологических процессов. Модели (1) – (2), (1) – (6) с различными целевыми функциями использовались для прогнозирования оборотов торгового эквайринга и пассивов физических лиц для 89 региональных филиалов крупного банка [16]. Модель (5) – (9) успешно используется нами в задачах медицинской диагностики [17], прогнозировании волатильности валютных курсов и показателей фондовых рынков [18]. Таким образом, приведенные в данной статье модели следует воспринимать как некоторый конструктор, на основе которого можно создать модель для решения задачи из конкретной предметной области.

Заключение. Предлагаемый нами подход обладает следующими достоинствами:

- 1) позволяет достигать качественных характеристик РП, сформулированных заказчиком, если это возможно, и предлагает оптимальную коррекцию пожеланий в противном случае;
- 2) позволяет минимизировать число признаков, необходимых для выполнения первого пункта;
- 3) позволяет давать геометрическую интерпретацию найденных решений, на основе которой практические специалисты могут переходить к содержательной интерпретации.

Математические модели достаточно простые, хорошо воспринимаются практическими специалистами, не требуют знания методов МП и навыков программирования на языках высокого уровня. Результаты моделей машинного обучения на базе математического программирования легко в дальнейшем использовать в моделях исследования операций.

Предлагаемый подход можно использовать в различных предметных областях.

Список источников

1. Вапник В.Н. Восстановление зависимостей по эмпирическим данным / В.Н. Вапник. – М.: Наука, 1979. – 447 с.
2. Bradley P., Fayyad U., Mangasarian O. Mathematical programming for data mining formulation and challenges. *Journal on Computing*, vol. 11, 1999, pp. 217-238.
3. Bennett K., Demiriz A., Taylor J. Linear programming boosting via column generation. *Proceedings of Seventeenth Conference on Machine Learning*, 2000, pp. 65-72.
4. Воронцов К.В. Лекции по методам оценивания и выбора моделей. – URL: <http://www.machinelearning.ru/wiki/images/2/2d/Voron-MI-Modeling.pdf> (дата обращения: 01.12.2022)
5. Wang B.Q., Chukova S., Lai C.D. On the relationship between regression analysis and mathematical programming. *Journal of applied mathematics and decision sciences*, vol. 8(2), 2004, pp. 131-140.
6. Yang L., Liu S., Tsoka S., Papageorgiou L.G. A regression tree approach using mathematical programming. *Expert Systems with Applications*, vol. 78, 2017, pp. 347-357.

7. Базилевский М.П. Сведение задачи отбора информативных регрессоров при оценивании линейной регрессионной модели по методу наименьших квадратов к задаче частично-булевого линейного программирования / М.П. Базилевский // Моделирование, оптимизация и информационные технологии, Воронеж, 2018. – Т. 6. – № 1 (20). – С. 108-117.
8. Горелик В.А. Решение задачи линейной регрессии с использованием методов матричной коррекции в метрике l_1 / В.А. Горелик, О.С. Трембачева // Журнал вычислительной математики и математической физики, 2016. – Т. 56. – № 2. – С. 202-207.
9. Мазуров В.Д. Комитеты системы линейных неравенств / В.Д. Мазуров, М.Ю. Хачай // Автоматика и телемеханика, 2004. – № 2. – С. 43-54.
10. Мазуров В.Д. Экзистенциальные вопросы комитетных конструкций / В.Д. Мазуров // Часть II. Вестник Южно-Уральского государственного университета, 2019. – Т. 19. – № 1. – С. 114-120.
11. Чернавин П.Ф. Анализ и прогнозирование выхода годного и прочности агломерата на основе моделей математического программирования / П.Ф. Чернавин, А.В. Малыгин, Т.В. Деткова [и др.] // Черные Металлы, 2021. – № 12. – С. 20-24.
12. Чернавин П.Ф. Машинное обучение на основе задач математического программирования / П.Ф. Чернавин, Д.Н. Гайнанов, В.Н. Панкращенко [и др.] – М.: Наука, 2021. – 128 с.
13. Чернавин П.Ф. Имитационная модель подбора технологических параметров для получения агломерата с высокими потребительскими свойствами на основе метода комитетов / П.Ф. Чернавин, А.Ф. Малыгин, Т.В. Деткова [и др.] // Черные металлы, 2022. – № 3. – С. 10-14.
14. Ащепков Л.Т. О построении максимального куба, вписанного в заданную область. / Л.Т. Ащепков // Журнал вычислительной математики и математической физики, 1980. – Т. 20. – № 2. – С. 510-513.
15. Meindl B., Templ M. Analysis of commercial and free and open source solvers for linear optimization problems. Institut f. Statistik u. Wahrscheinlichkeitstheorie, Austria, 2012, available at: https://www.researchgate.net/publication/265117825_Analysis_of_commercial_and_free_and_open_source_solvers_for_linear_optimization_problems (accessed: 12/01/2022)
16. Чернавин П.Ф. Сведение задач регрессии к задачам линейного целочисленного программирования. Анализ, моделирование, управление, развитие социально-экономических систем / П.Ф. Чернавин, Ф.П. Чернавин, Н.П. Чернавин // Сборник научных трудов XIV Всероссийской с международным участием школы-симпозиума АМУР-2020, Симферополь, 2020. – С. 383-386
17. Тюлькова Т.Е. Практическое применение различных методов машинного обучения на примере активности туберкулезного процесса у лиц с минимальными туберкулезными изменениями, выявленными на рентгенограмме органов грудной клетки. / Т.Е. Тюлькова, П.Ф. Чернавин, Н.П. Чернавин [и др.] // Клинический вестник им. А.И. Бурназяна, 2022. – № 2. – С. 64-73.
18. Чернавин Н.П. Прогнозирование волатильности курса валют методом комитетов. / Н.П. Чернавин // Вестник Челябинского государственного университета, 2019. – № 11(43). – С. 82-94.

Чернавин Павел Федорович. к.э.н., доцент кафедры Аналитика больших данных и методы видеоанализа, Уральский федеральный университет. Научные интересы: построение математических моделей с использованием методов исследования операций и машинного обучения. AuthorID: 117430, SPIN:6370-8103, ORCID: 0000-0003-3214-3906, chernavin.p.f@gmail.com, 620002, Свердловская область, г. Екатеринбург, ул. Мира д. 19.

Чернавин Николай Павлович. ассистент кафедры Аналитика больших данных и методы видеоанализа, Уральский федеральный университет, AuthorID : 971565, SPIN:5722-9436, ORCID: 0000-0003-3214-3906, ch_k@mail.ru, 620002, Свердловская область, г. Екатеринбург, ул. Мира д. 19.

Чернавин Федор Павлович. к.э.н., доцент кафедры Моделирование управляемых систем, Уральский федеральный университет, AuthorID: 971595, SPIN: 9237-5190, ORCID: 0000-0003-3214-3906, chernavin_fedor@mail.ru, 620002, Свердловская область, г. Екатеринбург, ул. Мира д. 19.

UDC: 004.855.5

DOI:10.25729/ESI.2023.30.2.005

Optimization models for selecting parameters of technological processes based on the machine learning results

Pavel F. Chernavin, Nikolai P. Chernavin, Fedor P. Chernavin

Ural Federal University, Russia, Yekaterinburg, *p.f.chernavin@urfu.ru*

Abstract. When solving practical problems, quite often there is a need for the simultaneous application of machine learning and operations research methods. Since many methods for solving problems in both areas can be based on fundamentally different mathematical tools, it will be impossible to combine their results into a single model. This article presents an interconnected set of machine learning and operations research models designed to select the parameters of the technological processes. All models have a common mathematical formulation based on the mathematical programming problems with the partial-integer variables. The models have been tested on a real problem of selecting the composition of the charge and technological parameters of sinter production. It is presented in the sequence of their occurrence in the process of solving the problems set by the customers of the research. The first stage is based on solution of the regression problems with a selection of the most informative features and the degree of their influence on the output features is carried out. Then, based on the classification problems, the recommended areas of controlled input features are determined to obtain the high-quality products. These areas can have a rather complex geometric configuration in a feature space. Further, within the framework of the operations research problems, the reference states of a process are determined, to which it is necessary to strive. At the final stage, the results of all previous studies are combined into a single optimization model, which can be supplemented with the results of the researches obtained from other sources of information, if these results can be represented as the linear constraints. The proposed approach to the parameter optimization can be used in the various subject areas.

Keywords: mathematical programming, machine learning, operations research, regression, classification, strength of sinter

References

1. Vapnik V. N. Vosstanovlenie zavisimostej po empiricheskim dannym [Recovery of dependences on empirical data]. Moscow, Nauka, 1979, 447 p.
2. Bradley P., Fayyad U., Mangasarian O. Mathematical programming for data mining formulation and challenges. *Journal on Computing*, v. 11, 1999, pp. 217-238.
3. Bennett K., Demiriz A., Taylor J. Linear programming boosting via column generation. *Proceedings of Seventeenth Conference on Machine Learning*, 2000, pp. 65-72.
4. Voroncov K.V. Lekcii po metodam ocenivaniya i vybora modelej [Lectures on estimation methods and model selection], available at: <http://www.machinelearning.ru/wiki/images/2/2d/Voron-MI-Modeling.pdf> (accessed: 01/12/2022).
5. Wang B.Q., Chukova S., Lai C.D. On the relationship between regression analysis and mathematical programming. *Journal of applied mathematics and decision sciences*, vol. 8(2), 2004, pp. 131-140.
6. Yang L., Liu S., Tsoka, S., Papageorgiou L.G. A regression tree approach using mathematical programming. *Expert systems with applications*, vol. 78, 2017, pp. 347-357.
7. Bazilevskij M.P. Svedenie zadachi otbora informativnyh regressorov pri ocenivanii linejnoj regressionnoj modeli po metodu naimen'shix kvadratov k zadache chastichno-bulevogo linejnogo programmirovaniya [Reduction of the problem of selecting informative regressors when evaluating a linear regression model using the least squares method to the problem of partially Boolean linear programming]. *Modelirovaniye, optimizatsiya i informatsionnyye tekhnologii [Modeling, optimization and information technologies]*, Voronezh, 2018, v.6, no.1 (20), pp. 108-117.
8. Gorelik V.A., Trembacheva O.S. Reshenie zadachi linejnoj regressii s ispol'zovaniem metodov matrichnoj korektsii v metrike [Solving a linear regression problem using matrix correction methods in metrics]. *Zhurnal vychislitel'noy matematiki i matematicheskoy fiziki [Journal of computational mathematics and mathematical physics]*, no. 2, 2016, pp. 202-207.
9. Mazurov V.D., Khachai M.Yu. Komitety sistemy linejnyh neravenstv [Committees of the system of linear inequalities]. *Avtomatika i telemekhanika [Automation and telemechanics]*, no.2, 2004, pp. 43-54.
10. Mazurov V.D. Ekzistencial'nye voprosy komitetnyh konstrukcij [Existential questions of committee constructions]. *Vestnik Yuzhno-Ural'skogo gosudarstvennogo universiteta [Part II. Bulletin of the South Ural state university]*, 2019, vol.19, no.1, pp. 114-120.

11. Chernavin P.F., Malygin A.V., Detkova T.V. [et al.]. Analiz i prognozirovaniye vyhoda godnogo i prochnosti aglomerata na osnove modelej matematicheskogo programmirovaniya [Analysis and prediction of yield and strength of sinter based on mathematical programming models]. Chernyye metally [Black metals], no. 12, 2021, pp. 20-24.
12. Chernavin P.F., Gajnanov D.N., Pankrashchenko V.N. [et al.]. Mashinnoe obuchenie na osnove zadach matematicheskogo programmirovaniya [Machine learning based on mathematical programming problems]. Moscow, Nauka [Moscow, Science], 2021, 128 p.
13. Chernavin P.F., Malygin A.V., Detkova T.V. [et al.]. Imitacionnaya model' podbora tekhnologicheskikh parametrov dlya polucheniya aglomerata s vysokimi potrebitel'skimi svoystvami na osnove metoda komitetov [Simulation model of selection of technological parameters for obtaining agglomerate with high consumer properties based on the method of committees]. Chernyye metally [Black metals], no. 3, 2022, pp. 10-14.
14. Ashchepkov L.T. O postroyenii maksimal'nogo kuba, vpisannogo v zadannuyu oblast'. [On the construction of a maximum cube inscribed in a given area]. Zhurnal vychislitel'noy matematiki i matematicheskoy fiziki [Journal of computational mathematics and mathematical physics], 1980, vol.20, no. 2., pp. 510-513.
15. Meindl B., Templ M. Analysis of commercial and free and open source solvers for linear optimization problems. Institut f. Statistik u. Wahrscheinlichkeitstheorie, Austria, 2012, available at: https://www.researchgate.net/publication/265117825_Analysis_of_commercial_and_free_and_open_source_sol. (accessed: 12/01/2022).
16. Chernavin P.F., Chernavin F.P., Chernavin N.P. Svedenie zadach regressii k zadacham linejnogo celochislennogo programmirovaniya. Analiz, modelirovaniye, upravleniye, razvitiye social'no-ekonomicheskikh sistem [Reduction of regression problems to linear integer programming problems. Analysis, modeling, management, development of socio-economic systems]. Sbornik nauchnykh trudov XIV Vserossiyskoy s mezhdunarodnym uchastiyem shkoly-simpoziuma AMUR-2020 [Collection of scientific papers of the XIV All-Russian School-symposium with international participation AMUR-2020], Simfiropol, 2020, pp. 383-386.
17. Tyulkova T.E., Chernavin P.F., Chernavin N.P. et. al. Prakticheskoe primeneniye razlichnykh metodov mashinnogo obucheniya na primere aktivnosti tuberkuleznogo processa u lic s minimal'nymi tuberkuleznymi izmeneniyami, vyyavlenymi na rentgenogramme organov grudnoj kletki. [Practical application of various machine learning methods on the example of the activity of the tuberculosis process in persons with minimal tuberculosis changes detected on the chest X-ray.] Klinicheskiy vestnik im. A.I. Burnazyana [A.I. Burnazyan clinical bulletin], no. 2, 2022, pp. 64-73.
18. Chernavin N.P. Prognozirovaniye volatil'nosti kursa valyut metodom komitetov. Vestnik Chelyabinskogo gosudarstvennogo universiteta. [Forecasting the volatility of the exchange rate by the method of committees.] Vestnik Chelyabinskogo gosudarstvennogo universiteta [Bulletin of chelyabinsk state university], 2019, no. 11(43), pp. 82-94.

Chernavin Pavel Fedorovich. Ph.D., Associate Professor of the Department of Big Data Analytics and Video Analysis Methods, Ural Federal University. Research interests: building mathematical models using operations research and machine learning methods. AuthorID: 117430, SPIN:6370-8103, ORCID: 0000-0003-3214-3906, chernavin.p.f@gmail.com, 620002, Russia, Sverdlovsk region, Ekaterinburg, st. Mira 19.

Chernavin Nikolay Pavlovich. Assistant Professor of Big Data Analytics and Video Analysis Methods, Ural Federal University, AuthorID: 971565, SPIN:5722-9436, ORCID: 0000-0003-3214-3906, ch_k@mail.ru, 620002, Russia, Sverdlovsk region, Ekaterinburg, st. Mira 19.

Chernavin Fedor Pavlovich. Ph.D., Associate Professor, Department of Controlled Systems Modeling, Ural Federal University, AuthorID: 971595, SPIN: 9237-5190, ORCID: 0000-0003-3214-3906, chernavin_fedor@mail.ru, 620002, Russia, Sverdlovsk region, Ekaterinburg, st. Mira 19.

Статья поступила в редакцию 10.01.2023; одобрена после рецензирования 28.04.2023; принята к публикации 04.05.2023.

The article was submitted 01/10/2023; approved after reviewing 04/28/2023; accepted for publication 05/04/2023.