DOI 10.24412/2221-2574-2022-4-69-78 УДК 519.711.2

РАЗРАБОТКА МАТЕМАТИЧЕСКОЙ МОДЕЛИ И АЛГОРИТМОВ ДЛЯ СИСТЕМЫ ПОСТРОЕНИЯ АППРОКСИМИРУЮЩЕГО ПОЛИНОМА

Киселёва Владислава Андреевна
магистрант Муромского института (филиала) ФГБОУ ВО «Владимирский государственный университет имени Александра Григорьевича и Николая Григорьевича Столетовых».

Рыжкова Мария Николаевна
кандидат технических наук, доцент кафедры физики и прикладной математики Муромского института (филиала) ФГБОУ ВО «Владимирский государственный университет имени Александра Григорьевича и Николая Григорьевича Столетовых».
E-mail: masmash@mail.ru

Адрес: 602264, Российская Федерация, Владимирская обл., г. Муром, ул. Орловская, д. 23.

Аннотация: В статье рассмотрен процесс разработки математической модели системы построения аппроксимирующего полинома. В качестве исходной модели для построения уравнения был выбран полином пятой степени, который может содержать от 1 до 5 нецелевых факторов. В работе была построена функциональная модель системы, которая содержит в себе такие блоки как: блок предварительной обработки данных; блок отбора признаков при построении уравнения; блок построения аппроксимирующего уравнения на основе метода наименьших квадратов; блок тестирования, который определяет точность построения модели. В основу работы каждого блока были положены алгоритмы, использующие известные методы анализа экспериментальных данных — метод наименьших квадратов и методы корреляционного и регрессионного анализа.

Ключевые слова: математическая модель, функциональная схема, аппроксимация данных, алгоритмы аппроксимации, полином регрессии, аппроксимирующий полином.

Введение

Аналитическое представление экспериментальных данных необходимо и важно, особенно для обеспечения высокой точности проводимых исследований. Аппроксимация данных является одной из задач регрессионного анализа, который стремительно развивается в машинном обучении. Преимущество данного подхода заключается в получении единственной модели, которая полностью опишет множественные взаимные связи между входными и выходными данными, даже в случае нелинейности связи между ними. Аппроксимация распространена в различных областях науки и техники, например, для прогнозирования величин или для фильтрации помех.

Целью научного исследования является разработка модели и алгоритмов работы системы построения аппроксимирующего полинома для обработки экспериментальных данных.

В качестве исходной модели для построения уравнения выберем полином пятой степени, который может содержать от 1 до 5 нецелевых факторов. Такая степень достаточна для полного описания закономерностей поведения экспериментальных данных, увеличение степени чаще всего не приводит к увеличению точности аппроксимации, зато на порядок увеличивает сложность построения модели.

Для достижения поставленной цели, требуется выполнить следующие задачи:

– изучить предметную область исследования;

– выбрать алгоритмы предварительной обработки данных;

– выбрать алгоритм построения аппроксимирующего полинома;

– описать способ вычисления оценки точности аппроксимации;

– разработать функциональную модель системы.

В ходе разработки системы используются известные методы обработки экспериментальных

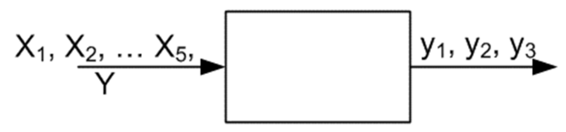

Рис. 1. Модель «Чёрный ящик»

ных данных, описанные в работах [1–10].

Разработка функциональной модели

При разработке информационной системы начинают в первую очередь с разработки её модели и алгоритмов работы. Опишем разработку функциональной модели системы, которая подразумевает математическое описание способов преобразования входной информации системы в выходную [11].

Для реализации функциональной модели необходимо последовательно разработать:

1) модель «чёрный ящик» для определения вида входных и выходных параметров;

2) модель состава для определения основных блоков обработки данных;

3) структурную модель для определения перемещения данных в блоках;

4) функциональную модель для определения работы отдельных блоков системы.

Модель типа «черный ящик» (рис. 1) предполагает описание взаимодействия системы с пользователем. Для этого в модели показаны данные, которые поступают на входы системы, и выходные данные, которые выводятся пользователю в качестве результата работы. На вход системы подаётся набор экспериментальных данных, который может содержать 1-5 нецелевых факторов *X* и одно значение отклика *Y*.

Выходными данными системы являются: y1 — аппроксимирующее уравнение в виде полинома 5-ой степени; y2 — значение точности работы системы в процентах; y3 — рекомендация для пользователя.

На рис. 2 представлена модель состава системы, которая включает в себя основные блоки обработки экспериментальных данных.

На рис. 3 основное внимание уделено перемещению данных внутри системы. Таким образом, получаем модель структуры, которая содержит не только основные блоки, но и взаимосвязи между ними.

Перейдём к математическому описанию преобразования данных, в результате чего получим функциональную модель системы (рис. 4).

Опишем каждый из блоков полученной функциональной модели.

K_1 — блок предварительной обработки данных:

- K_{11} — разбиение входного набора на обучающую (80%) и тестовую (20%) выборки;
- K_{12} — предварительная обработка данных:

  - K_{12_1} — нормирование обучающей выборки;
  - K_{12_2} — восстановление пропущенных значений;
  - K_{12_3} — поиск аномалий и замена их на медианное значение.

K_2 — блок отбора признаков:

- K_{21} — вычисление матрицы корреляции;
- K_{22} — поиск линейной зависимости между нецелевыми факторами и откликом.

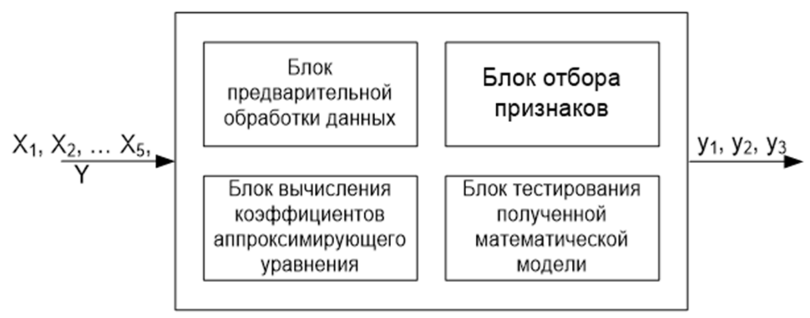

Рис. 2. Модель состава

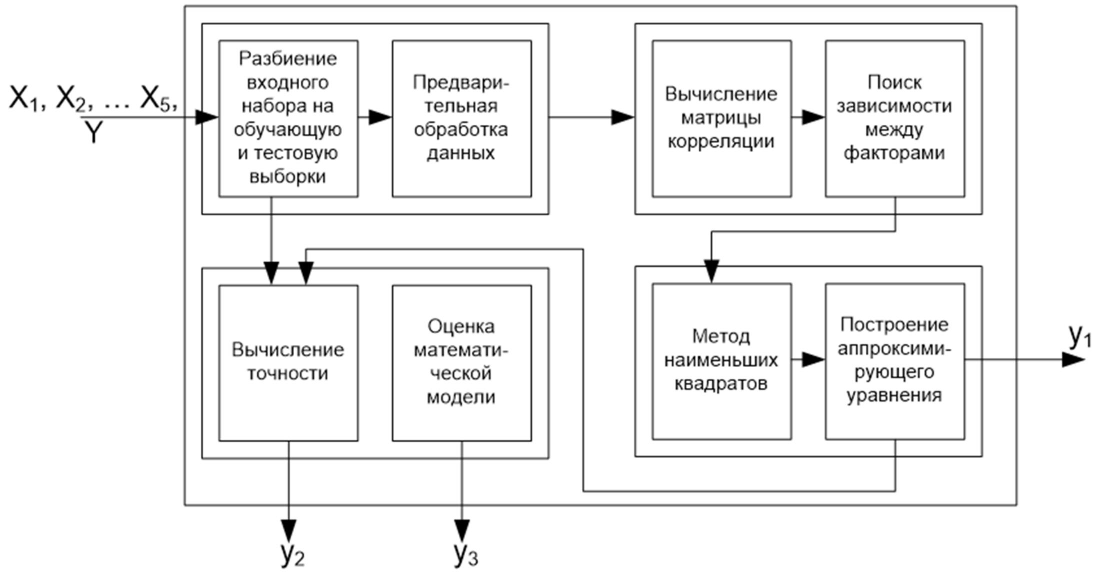

Рис. 3. Модель структуры

K_3 — блок вычисления коэффициентов аппроксимирующего уравнения:

- K_{31} — применение метода наименьших квадратов для вычисления коэффициентов аппроксимирующего уравнения:
    - K_{31_1} — составление системы уравнений по предварительно обработанным данным;
    - K_{31_2} — нормирование системы уравнений;
    - K_{31_3} — решение системы методом Гаусса;
- K_{32} — переход от нормированных коэффициентов и вычисление свободного члена и составление аппроксимирующего уравнения.

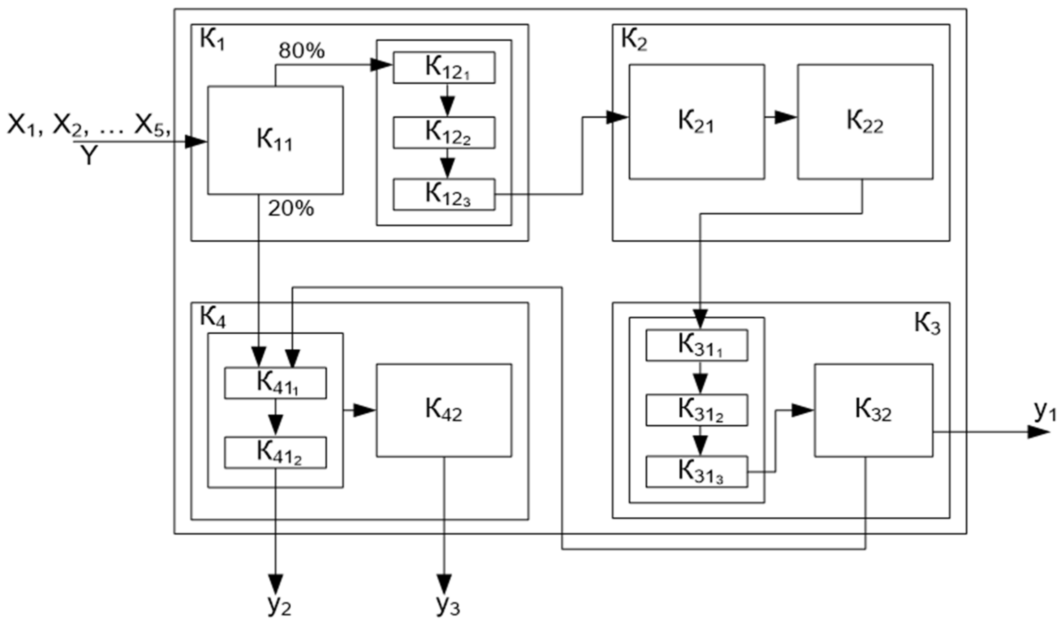

Рис. 4. Функциональная модель

K_4 — блок тестирования, полученной математической модели:

- K_{41} — вычисление точности, которое заключается в подстановке нецелевых факторов из тестовой выборки в полученную математическую модель для вычисления теоретического значения отклика, которое после сравнивается с экспериментальным значением отклика;
- K_{42} — оценка математической модели, а именно вывод рекомендации для пользователя если значение точности менее 90%.

Описание алгоритмов обработки данных в системе

Среди алгоритмов работы системы выделим наиболее важные:

- алгоритмы предварительной обработки данных, которые анализируют исходные данные, поступающие в систему, нормализуют полученные значения, восстанавливают пропущенные значения, если таковые присутствуют в выборке, находят выбросы и аномалии в значениях выборки и корректируют их;
- алгоритмы отбора признаков при построении уравнения используют методы корреляционного анализа для установления значимости факторов при построении функциональной зависимости;
- алгоритм построения аппроксимирующего уравнения на основе метода наименьших квадратов;
- алгоритм тестирования полученной модели, который определяет точность построения модели с помощью одной из простейших характеристик *MAPE*.

1. Блок предварительной обработки данных K_1 выполняет предварительный отбор и предобработку данных. Эта процедура необходима в силу высокой чувствительности регрессионной модели к вариациям входного набора данных.

В подблоке K_{11} генеральная совокупность делится на обучающую и тестовую выборки, 80% данных передаются в подблок K_{12}, а 20% в блок тестирования модели K_4.

В подблоке K_{12} обучающая выборка предварительно обрабатывается.

K_{12_1} — нормирование обучающей выборки.

Для нормирования данных стоит выполнить преобразование:

$$p_i^{'} = \frac{p_i - \bar{p}}{s}, \qquad (1)$$

где $\bar{p}$ — среднее значение признака; s — среднее квадратичное отклонение.

Среднее квадратичное отклонение вычисляется по формуле:

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(p_i - \bar{p})^2}, \qquad (2)$$

где n — количество значений признака.

K_{12_2} — восстановление пропущенных значений.

Пропуски могут возникать по разным причинам, и есть различные решения обхода данной проблемы. Можно заменить пустое значение на «0», моду или медиану фактора, а также можно избавиться от проблемы путём удаления строки или столбца, в котором был найден пропуск. Данные подходы могут значительно повлиять на качество аппроксимации, поэтому стоит применить более действенный метод.

Для вычисления пропущенного значения нужно взять линейную комбинацию со значениями признака, тогда формула для восстановления данных выглядит так:

$$P(A) = \frac{1}{\sum_{i=1}^{n}\frac{1}{\rho(A, A_j)}}\sum_{i=1}^{n}\frac{P(A_j)}{\rho(A, A_j)}, \qquad (3)$$

где $\rho(A, A_j)$ — расстояние от объекта с пропущенным значением до другого объекта; $P(A_j)$ — значение признака объекта.

Расстояние между объектами можно вычислить при помощи максимальной метрики:

$$\rho(P,Q) = \max\{|p_1 - q_1|, |p_2 - q_2|, \ldots |p_n - q_n|\}. \qquad (4)$$

$К_{12_3}$ — поиск аномалий и замена их на медианное значение.

Выбросы — это измерения, лежащие аномально далеко от результатов общей выборки. Подобные явления негативно влияют на результат аппроксимации. Для определения более точной функциональной зависимости между данными необходимо исключить выбросы из выборки.

Первым шагом в данном анализе является упорядочивание исходного набора данных по возрастанию, и вычисление в нём медианы (значение, делящее выборку пополам). Пусть имеется ряд значений с количеством элементов n, тогда в случае нечётного количества элементов медианой является значение, стоящее ровно в середине выборки:

$$M_e = x_{\frac{n+1}{2}}. \tag{5}$$

Если выборка содержит чётное количество элементов, медиана вычисляется по формуле:

$$M_e = \frac{x_{\frac{n+1}{2}} + x_{\frac{n-1}{2}}}{2}. \tag{6}$$

Верхний Q_1 и нижний Q_2 квартили вычисляются подобно медиане. Для этого выборка делится пополам и находятся медианы полученных рядов. Диапазон между квартилями вычисляется как:

$$D = Q_2 - Q_1. \tag{7}$$

Важнейший этап нахождения выбросов, это определение внешних границ выборки. Внешние границы — это отрезок, который включает все значения ряда, не являющиеся выбросами, координаты которого вычисляются по формулам:

$$G_1 = Q_2 - 3D, \tag{8}$$

$$G_2 = Q_2 + 3D. \tag{9}$$

Все значения из набора данных, которые не входят в интервал $[G_1, G_2]$, считаются выбросами.

Таким образом, алгоритм определения выбросов в данных:

- упорядочить данные по возрастанию;
- вычислить медиану упорядоченного ряда;
- найти нижний квартиль упорядоченного ряда;
- найти верхний квартиль упорядоченного ряда;
- вычислить диапазон между квартилями;
- найти внешние границы набора данных;
- выявить выбросы в данных;
- заменить выбросы на медиану данного ряда;
- повторить алгоритм для каждого фактора.

2. Блок отбора признаков.

Одной из важнейших задач обработки данных, является анализ зависимостей между анализируемыми величинами. Для многих явлений характерно, что существующие связи «размыты» под влиянием случайных факторов. Для установления взаимосвязи между факторами и применяют корреляционный анализ.

$К_{21}$ — вычисление матрицы корреляции.

Для вычисления выборочного коэффициента корреляции необходимо знать ковариацию и дисперсию выбранных признаков:

$$\operatorname{cov}_в(X,Y) = \overline{XY} - \overline{X}\,\overline{Y}, \tag{10}$$

где $\operatorname{cov}_в(X,Y)$ — выборочная ковариация между факторами X и Y; X — нецелевой признак; Y — результативный признак;

$$D_в(X) = \overline{X^2} - \overline{X}^2, \tag{11}$$

где $D_в(X)$ — дисперсия выбранного фактора.

Тогда коэффициент корреляции между выбранными факторами:

$$r_в(X,Y) = \frac{\operatorname{cov}_в(X,Y)}{\sqrt{D_в(X)\cdot D_в(Y)}}. \tag{12}$$

Матрица корреляции для 5-ти нецелевых и 1-го целевого признаков:

$$R = \begin{pmatrix} 1 & r_в(X_2,X_1) & \dots & r_в(X_1,Y) \\ r_в(X_1,X_2) & 1 & \dots & r_в(X_2,Y) \\ r_в(X_1,X_3) & r_в(X_2,X_3) & \dots & r_в(X_3,Y) \\ r_в(X_1,X_4) & r_в(X_2,X_4) & \dots & r_в(X_4,Y) \\ r_в(X_1,X_5) & r_в(X_2,X_5) & \dots & r_в(X_5,Y) \\ r_в(X_1,Y) & r_в(X_2,Y) & \dots & 1 \end{pmatrix}. \tag{13}$$

K_{22} — поиск линейной зависимости между нецелевыми факторами и откликом.

Коэффициент корреляции определяет количественную оценку тесноты функциональной связи. Будем определять коэффициент корреляции как между откликом и нецелевыми факторами, так и между самими нецелевыми факторами. Это позволит убрать из дальнейшего исследования те нецелевые факторы, связь между которыми линейна, а также те факторы, которые не влияют на отклик.

Факторы, которые необходимо исключить:

- если $r_в(X_n, X_m) > 0,7$ то один из нецелевых факторов нужно исключить;

- если $r_в(X_n, Y) < 0,3$, то нецелевой фактор нужно исключить.

3. Блок составления аппроксимирующего уравнения

После того, как обучающая выборка предварительно обработана, а факторы не имеющие зависимости исключены, можно приступать к процедуре составления аппроксимирующего полинома.

Метод регрессионного анализа использует описание объекта исследования в виде некоторого полинома — отрезка ряда Тейлора, в который разлагается неизвестное уравнение связи отклика объекта Y и входных факторов X. Необходимо использовать полином такой формы, который включает в себя все возможные сочетания факторов:

$$\begin{aligned}\varphi(x_1, x_2, \dots x_k) = \beta_0 + \sum_{i=1}^{k}\beta_i x_i + \sum_{i=1, j>i}^{k}\beta_{ij}x_i x_j + \\ + \sum_{i=1, j>i, q>j}^{k}\beta_{ijq}x_i x_j x_q + \dots + \\ + \sum_{i=1}^{k}\beta_{ii}x_i^2 + \sum_{i=1}^{k}\beta_{iii}x_i^3 + \dots + \\ + \sum_{i=1}^{k}\beta_{iiiii}x_i^5,\end{aligned} \tag{14}$$

где $\varphi(x_1, x_2, \dots x_k)$ — отражение неизвестной функции истинного отклика; $x_1, x_2, \dots x_k$ — входные факторы; β — коэффициенты, являющиеся производными вида $\frac{\partial\varphi}{\partial x_i}$; k — количество входных факторов; i, j, q — номера входных факторов.

Число входных факторов может быть от одного до пяти для одного значения функции, тогда на переменную k накладывается ограничение $1 \le k \le 5, k \in Z$, где Z — множество целых чисел.

Так как по числу факторов математическая модель объекта исследования не всегда является полной, что делает отклик (соответствующее значение y для входных факторов) случайной величиной, поэтому зависимость $\varphi(x_1, x_2, \dots x_k)$ не даёт точной связи между y и x. В связи с этим β необходимо заменить выборочными коэффициентами, которые и необходимо найти. Тогда уравнение (14) примет вид:

$$\begin{aligned}y_g = b_0 + \sum_{i=1}^{k_1} b_i x_i + \sum_{i=1, j>i}^{k_2} b_{ij}x_i x_j + \\ + \sum_{i=1}^{k_3} b_{iii}x_i^3 + \dots + \sum_{i=1}^{k_5} b_{iiiii}x_i^5,\end{aligned} \tag{15}$$

где b — выборочные коэффициенты; y_g — значение отклика.

В подблоке K_{31} применяется метод наименьших квадратов для вычисления коэффициентов аппроксимирующего уравнения.

K_{31_1} — составление системы уравнений по предварительно обработанным данным.

Чтобы приступить к обработке по методу наименьших квадратов необходимо составить систему уравнений, число строк в которой будет соответствовать числу строк обучающей выборки:

$$\begin{cases} b_1 x_{11} + \dots + b_k x_{k1} + \dots + b_{11}x_{11}^2 + \dots + b_{kkkkk}x_{k1}^5 = y_1, \\ b_1 x_{12} + \dots + b_k x_{k2} + \dots + b_{12}x_{12}^2 + \dots + b_{kkkkk}x_{k2}^5 = y_2, \\ \dots \\ b_1 x_{1g} + \dots + b_k x_{kg} + \dots + b_{1g}x_{1g}^2 + \dots + b_{kkkkk}x_{kg}^5 = y_g, \end{cases} \tag{16}$$

где g — это номер строки обучающей выборки.

$К_{31_2}$ — нормирование системы уравнений.

Процедура преобразования системы в нормальный вид состоит из двух шагов. На первом необходимо каждое уравнение из системы (16) умножить на множитель при коэффициенте b_0 (в данном случае множитель равен 1). Второй шаг заключается в суммировании каждого уравнения полученной системы сверху вниз. В итоге получим одно уравнение, которое будет являться первым для системы нормальных уравнений:

$$\begin{aligned}\sum x_1 y = b_1 \sum x_1^2 + \ldots + b_k \sum x_1 x_k + \ldots + \\ + b_{12} \sum x_1^2 x_2 + \ldots + b_{kk} \sum x_1 x_k^2 + \ldots + \\ + b_{11111} \sum x_1^6 + \ldots + b_{kkkkk}.\end{aligned} \quad (17)$$

Следующим шагом станет умножение каждого уравнения исходной системы (16) на множитель при b_1 (первое уравнение системы умножаем на x_{11}, второе на x_{12}, g-ое на x_{1g}). Снова суммируем полученные уравнения сверху вниз, получая при этом второе уравнение системы нормальных уравнений. Процедура продолжается для всех b, тем самым формируются остальные уравнения нормальной системы. По итогу в сформированной системе число уравнений соответствует числу коэффициентов в исходном полиноме.

Система нормальных уравнений для полинома, которая совместна и минимизирует отклонения, будет иметь следующий вид:

$$\begin{cases}\sum x_1 y = b_1 \sum x_1^2 + \ldots + b_k \sum x_1 x_k + \ldots + \\ \quad + b_{12} \sum x_1^2 x_2 + \ldots + b_{kk} \sum x_1 x_k^2 + \ldots + \\ \quad + b_{11111} \sum x_1^6 + \ldots + b_{kkkkk} \sum x_1 x_k^5, \\ \sum x_2 y = b_1 \sum x_1 x_2 + \ldots + b_k \sum x_2 x_k + \ldots + \\ \quad + b_{12} \sum x_1 x_2^2 + \ldots + b_{kk} \sum x_2 x_k^2 + \ldots + \\ \quad + b_{11111} \sum x_1^5 x_2 + \ldots + b_{kkkkk} \sum x_2 x_k^5, \\ \quad \ldots \\ \sum x_k^5 y = b_1 \sum x_1 x_k + \ldots + b_k \sum x_k^6 + \ldots + \\ \quad + b_{12} \sum x_1 x_2 x_k^5 + \ldots + b_{kk} \sum x_k^7 + \ldots + \\ \quad + b_{11111} \sum x_1^5 x_k^5 + \ldots + b_{kkkkk} \sum x_k^{10}.\end{cases} \quad (18)$$

$К_{31_3}$ – решение системы методом Гаусса.

Нормированная система является линейной, следовательно, её можно решить методом Гаусса, который заключается в последовательном исключении неизвестных из уравнений системы. Для решения системы (18) необходимо найти расширенную матрицу.

Правая часть системы (18) — это произведение матрицы на вектор коэффициентов b. Обозначим их как F и B соответственно и выделим, тогда:

$$F = \begin{pmatrix} \sum x_1^2 & \ldots & \sum x_1^2 x_2 & \ldots & \sum x_1 x_k^5 \\ \sum x_1 x_2 & \ldots & \sum x_1 x_2^2 & \ldots & \sum x_2 x_k^5, \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ \sum x_1 x_k & \ldots & \sum x_1 x_2 x_k^5 & \ldots & \sum x_k^{10}. \end{pmatrix} \quad (18)$$

и

$$B = \begin{pmatrix} b_1 \\ \ldots \\ b_{12} \\ \ldots \\ b_{kkkkk} \end{pmatrix}. \quad (20)$$

Левая часть системы (18) — это вектор суммы парных произведений:

$$Y = \begin{pmatrix} \sum x_1 y \\ \sum x_2 y \\ \ldots \\ \sum x_k^5 y \end{pmatrix}. \quad (21)$$

Тогда расширенная матрица для системы (18):

$$\overline{A} = (F \mid Y). \quad (22)$$

Для решения системы (18) нужно провести элементарные преобразования с матрицей $\overline{A}$:

- изменить порядок строк (что соответствует изменению порядка уравнений);
- умножить строки на любые отличные от нуля числа (что соответствует умножению соответствующих уравнений на эти числа);
- прибавить к любой строке расширенной матрицы любую другую её строку, умноженную на любое число (что соответствует прибавлению к одному из уравнений системы другого уравнения, умноженного на это число).

После выполнения элементарных преобразований получается система, равносильная исходной, матрица которой приведена к ступенчатому виду. Далее записывается новая система линейных уравнений, соответствующая матрице ступенчатого вида. Начиная с последнего уравнения, находится решение для данной системы. В результате решения получаем набор нормированных коэффициентов b, которые и являются коэффициентами полинома регрессии (15).

В подблоке K_{32} выполняется переход от нормированных коэффициентов и вычисление свободного члена. Для вычисления исходных коэффициентов аппроксимирующего уравнения необходимо вычислить среднее квадратичное отклонение по формуле (2) для каждого произведения факторов x при коэффициентах b в уравнении (15).

Формула перехода от нормированных коэффициентов:

$$b = \beta \frac{s_y}{s_{x_i}}. \tag{23}$$

Формула вычисления свободного члена:

$$\begin{gathered} b_0 = \overline{y} - b_1 \overline{x_1} - \ldots - b_k \overline{x_k} - \ldots - \\ - b_{11111} \overline{x_k^5} - \ldots - b_{kkkkk} \overline{x_k^5}. \end{gathered} \tag{24}$$

Далее полученные коэффициенты β подставляются в исходное уравнение (14) и таким образом, составляется аппроксимирующий полином, который выводится пользователю.

4. Блок тестирования полученной модели.

Блок K_4 отвечает за тестирование полученной математической модели. В подблоке K_{41} вычисляется средняя абсолютная ошибка в процентах.

K_{41_1} — вычисление теоретических значений y на основе подстановки нецелевых факторов из тестовой выборки в полученный аппроксимирующий полином.

K_{41_2} — вычисление средней абсолютной ошибки в процентах.

Оценка качества построенной математической модели осуществляется по формуле:

$$MAPE = \frac{1}{n} \sum_{i=1}^{n} \frac{|y_i - y_i'|}{|y_i|} \cdot 100\%, \tag{25}$$

где $MAPE$ — средняя относительная ошибка, выраженная в процентах; n — объём тестовой выборки; y_i — истинное значение целевого признака; y_i' — полученное значение целевого признака.

В подблоке K_{42} полученное значение $MAPE$ сравнивается с пороговым значением «10%», которое наиболее часто используется экспертами при оценке точности регрессионной модели. Если ошибка превышает этот уровень, то делается вывод о некорректном наборе данных.

Заключение

В результате исследований была разработана модель работы системы для построения аппроксимирующего полинома пятой степени для количества нецелевых факторов от 1 до 5.

При разработке модели системы были использованы известные методы обработки экспериментальных данных. При этом были получены алгоритмы для предварительной обработки, которые обеспечивают анализ исходных данных, поступающих в систему, нормализацию полученных значений, восстановление пропусков, если таковые присутствуют в выборке, выявление выбросов и аномалий в выборке и их коррекцию. Разработанные алгоритмы отбора признаков используют методы корреляционного анализа для установления значимости факторов при построении функциональной зависимости. В основу построения блока аппроксимирующего уравнения были положены метод наименьших квадратов и алгоритм тестирования полученной модели

Литература

1. *Бердышев В.И., Петрак Л.В.* Аппроксимация функций, сжатие численной информации, приложения. Екатеринбург: УрО РАН, 1999. 296 с.

2. *Стрижов В.В.* Методы индуктивного порождения регрессионных моделей. М.: ВЦ РАН. 2008

3. *Овчинников А.В., Красночуб Е.К., Бронштейн В.М.* Обработка экспериментальных данных

методом наименьших квадратов [Электронный ресурс] // Вестник СГАУ. 2009. №3. URL: https://cyberleninka.ru/article/n/obrabotka-eksperimentalnyh-dannyh-metodom-naimenshih-kvadratov-1 (дата обращения: 15.10.2022).

4. *Овчинников А.В., Красночуб Е.К., Бронштейн В.М.* Обработка экспериментальных данных методом наименьших квадратов [Электронный ресурс] // Вестник СГАУ. 2010. №2. URL: https://cyberleninka.ru/article/n/obrabotka-eksperimentalnyh-dannyh-metodom-naimenshih-kvadratov (дата обращения: 21.10.2022).

5. *Ивановский Р.И.* Обработка экспериментальных данных энергетических объектов как начальный этап их моделирования [Электронный ресурс] // Информатика, телекоммуникации и управление. 2010. №2 (97). URL: https://cyberleninka.ru/article/n/obrabotka-eksperimentalnyh-dannyh-energeticheskih-obektov-kak-nachalnyy-etap-ih-modelirovaniya (дата обращения: 20.10.2022).

6. *Усмонов М.* Метод наименьших квадратов [Электронный ресурс] // Science and Education. 2021. №7. URL: https://cyberleninka.ru/article/n/metod-naimenshih-kvadratov-1 (дата обращения: 30.10.2022).

7. *Лукьянов В.Д.* О построении интерполяционно-аппроксимационного многочлена // Наносистемы: физика, химия, математика. 2012. №6. С. 5–15.

11. *Рыжкова М.Н.* Универсальная модель образовательной системы: кибернетический подход // Вестник РГРТУ. 2015. №53. С. 99–105.

12. *Рыжкова М.Н., Кутарова Е.И.* Когнитивное моделирование результатов образовательной деятельности студентов радиотехнического направления подготовки // Радиотехнические и телекоммуникационные системы. 2016. №2. С. 79–86.

13. *Рыжкова М.Н., Кутарова Е.И.* Модель процесса обучения студентов радиотехнического направления подготовки дисциплинам естественнонаучного цикла // Радиотехнические и телекоммуникационные системы. 2017. №2. С. 68–73.

.

Поступила 30 октября 2022 г.

English

DEVELOPMENT OF MATHEMATICAL MODEL AND ALGORITHMS FOR POLYNOMIAL APPROXIMANT ARCHITECTURE

Vladislava Andreyevna Kiselyova — Master's Degree Student, Murom Institute (branch) "Vladimir State University named after A.G. and N.G. Stoletovs".

Maria Nikolayevna Ryzhkova — PhD, Associate Professor, Associate Professor of Department of Physics and Applied Mathematics, Murom Institute (branch) "Vladimir State University named after A.G. and N.G. Stoletovs".

E-mail: masmash@mail.ru

Address: 602264, Russian Federation, Vladimir region, Murom, Orlovskaya st., 23.

Abstract: The paper objective is to develop a model and operation algorithms for polynomial approximant architecture to process testing data. Quintic was chosen as the source model to form the equation, which can comprise from 1 to 5 off-target factors. Such a degree is sufficient for a detailed description of behaviour regularities for testing data, degree increase would not result most often in approximation accuracy improvement but it contributes to complexity of model development by times. The article describes development of system functional model, which implies a mathematical description of methods to transform system input data into the output one. To implement the functional model there were consecutively reviewed the following: 1) black box" model to define the type of input and output parameters; 2) composition model to define basic data processing units; 3) structural model to define data migration in units; 4) functional model to define operation of individual system units. The following was developed as part of constructing a functional model of the system: 1) data pre-processing algorithms that analyze initial input data for the system, normalize the obtained values, restore missing values, if there are any in selection, find outliers and anomalies in selection values and correct them; 2) when forming an equation feature selection algorithms use methods of correlation analysis to determine factors' significance in constructing a functional dependence; 3) algorithm to form an approximating equation based on the least-squares method; 4) testing algorithm for the obtained model, which defines model construction accuracy using one of the simplest MAPE properties. Model development precedes the development of information system for testing data processing, which is currently relevant and in demand.

Keywords: mathematical model, functional diagram, data approximation, approximation algorithms, regression polynomial, polynomial approximant.

References

1. *Berdyshev V.I., Petrak L.V.* Approximation of functions, compression of numerical information, applications. Ekaterinburg: Ural Branch of the Russian Academy of Sciences, 1999. 296 p.

2. *Strizhov V.V.* Methods of inductive generation of regression models. Moscow: VTs RAN. 2008

3. *Ovchinnikov A.V., Krasnochub E.K., Bronstein V.M.* Processing of experimental data by the least squares method [Electronic source]. Vestnik SSAU. 2009. No. 3. URL: https://cyberleninka.ru/article/n/obrabotka-eksperimentalnyh-dannyh-methodom-naimenshih-kvadratov-1 (Access date: 15.10.2022).

4. *Ovchinnikov A.V., Krasnochub E.K., Bronstein V.M.* Processing of experimental data by the least squares method [Electronic source]. Vestnik SSAU. 2010. No. 2. URL: https://cyberleninka.ru/article/n/obrabotka-eksperimentalnyh-dannyh-methodom-naimenshih-kvadratov (Access date: 21.10.2022).

5. *Ivanovsky R.I.* Processing of experimental data of power facilities as the initial stage of their modeling. Informatics, telecommunications and management [Electronic source]. 2010. No. 2 (97). URL: https://cyberleninka.ru/article/n/obrabotka-eksperimentalnyh-dannyh-energeticheskih-obektov-kak-nachalnyy-etap-ih-modelirovaniya (Access date: 20.10.2022).

6. *Makhsud Tulqin Ogli Usmonov.* Method of least squares. Science and Education [Electronic source]. 2021. No. 7. URL: https://cyberleninka.ru/article/n/method-naimenshih-kvadratov-1 (Access date 30.10.2022).

7. *Lukyanov V.D.* On the construction of an interpolation-approximation polynomial. Nanosystems: physics, chemistry, mathematics. 2012. No. 6. Pp. 5–15.

11. *Ryzhkova M.N.* Universal model of the educational system: a cybernetic approach. Bulletin of the Russian State Technical University. 2015. No. 53. Pp. 99–105.

12. *Ryzhkova M.N., Kutarova E.I.* Cognitive modeling of the results of educational activity of students of the radio engineering direction of training. Radio engineering and telecommunication systems. 2016. No.2. pp. 79-86.

13. *Ryzhkova M.N., Kutarova E.I.* Model of the process of teaching students of the radio engineering direction of training disciplines of the natural science cycle. Radioengineering and telecommunication systems. 2017. No. 2. Pp. 68–73.