

ОБОБЩЕННАЯ МОДЕЛЬ КОНТРОЛЯ ПРАВИЛЬНОСТИ КОДОВ ОБЩЕРОССИЙСКИХ КЛАССИФИКАТОРОВ ТЕХНИКО- ЭКОНОМИЧЕСКОЙ И СОЦИАЛЬНОЙ ИНФОРМАЦИИ

САКОВ А.А.,

доктор экономических наук, доцент, директор Департамента общероссийских классификаторов технико-экономической и социальной информации Российского научно-технического центра информации по стандартизации, метрологии и оценке соответствия (ФГУП «СТАНДАРТИНФОРМ»)

Статья посвящена обоснованию модели контроля правильности кодов, установленных в общероссийских классификаторах технико-экономической и социальной информации на основе использования теории множеств.

Ключевые слова: общероссийские классификаторы; коды; контрольное число; модуль; весовые коэффициенты; взвешенная сумма.

UDC 006.027

GENERALIZED CONTROL MODEL FOR CODE ACCURACY USED IN RUSSIAN CLASSIFICATIONS OF TECHNICAL, ECONOMICAL AND SOCIAL INFORMATION

SAKOV A.A.,

doctorate degree in economic sciences, associate professor, head of the Department of Russian classifiers of technical, economical and social information at Russian scientific and technical centre for information on standardization, metrology and conformity assessment (FGUP «STANDARTINFORM»)

The article is devoted to substantiating the generalized model for code accuracy used in Russian classifications of technical, economical and social information and providing mathematical proof using sets theory.

Keywords: Russian classifications, codes, control number, module, weighting coefficients, weighted sum.

Необходимость контроля правильности кодов, установленных в общероссийских классификаторах, возникает, как правило, при первоначальном и последующем вводе в автоматизированные базы и банки данных технико-

экономической и социальной информации, содержащей коды, установленные в общероссийских классификаторах, а также при разработке и типографском издании общероссийских классификаторов и изменений к ним. При этом появляющиеся ошибки в кодах связаны в основном с человеческим фактором.

Для контроля правильности кодов в соответствии с Правилами стандартизации ПР 50.1.024-2005 «Основные положения и порядок проведения работ по разработке, ведению и применению общероссийских классификаторов» [1] используется алгоритм, приведенный в следующей методике расчета контрольного числа.

Разрядам кода в общероссийском классификаторе, начиная со старшего разряда, присваивается набор весов, соответствующий натуральному ряду чисел от 1 до 10. Если разрядность кода больше 10, то набор весов повторяется. Далее каждая цифра кода умножается на вес разряда и вычисляется сумма полученных произведений.

Контрольное число для кода представляет собой остаток от деления полученной суммы на модуль «11». При этом контрольное число должно иметь один разряд, значение которого находится в пределах от 0 до 9.

Если получается остаток, равный 10, то для обеспечения одноразрядного контрольного числа проводится повторный расчет с применением последовательности весов, сдвинутой на два разряда влево (3, 4, 5, ...).

Если и в случае повторного расчета остаток от деления вновь сохраняется равным 10, то значение контрольного числа проставляется равным «0». Например, применительно к Общероссийскому классификатору продукции (ОКП), код в котором составляет шесть десятичных цифр, расчет контрольного числа осуществляется следующим образом.

Код ОКП	5	6	2	8	2	1
Вес разрядов	1	2	3	4	5	6

Сумма произведений, полученных путем умножения каждой цифры кода на вес разряда, равна: $5x_1 + 6x_2 + 2x_3 + 8x_4 + 2x_5 + 1x_6 = 71$ (1)

Остаток от деления полученной суммы на модуль «11» равен 5:

$$71 : 11 = 6 \text{ (5)} \quad (2)$$

Контрольное число для данного кода по модулю «11» равно 5.

Наиболее простым способом организации контроля правильности кода по данной методике является автоматическое обнаружение ошибок, заключающихся в несоответствии кода и контрольного числа к нему, с последующим их исправлением вручную.

Несмотря на то, что данная методика применяется уже много лет, по её содержанию постоянно возникают вопросы. Поэтому в данной статье обосновывается модель контроля правильности кодов, установленных в общероссийских классификаторах технико-экономической и социальной информации, базирующаяся на использовании теории множеств. Данная модель является отражением методики, приведенной в [1].

Выражения (1) и (2) в общем случае можно записать в виде

$$\sum_{i=1}^S d_i h_i - nm = b \quad (3)$$

В выражении (3) S – длина кода, $d_1 \dots d_s$ – цифры, соответствующие i -м разрядам кода, $h_1 \dots h_s$ – весовые коэффициенты, соответствующие i -м разрядам кода, n – модуль, m – целое число, включая ноль, b – остаток от деления взвешенной суммы цифр кода на модуль.

Совокупность кодов общероссийского классификатора (например, ОКП) может рассматриваться как некоторое множество целых чисел, которое может быть разделено на непересекающиеся между собой классы чисел, дающих при делении взвешенной суммы цифр кода на модуль n один и тот же остаток b .

Обозначим указанные классы через

$$A_0 \dots A_k \dots A_l \dots A_{n-1} \quad (4)$$

Здесь для A_0 $b=0$, для A_k $b=k$, для A_l $b=l$, для A_{n-1} $b=n-1$.

Величины d_i , h_i , n , m , b из (3) также будут принадлежать к вполне определенным классам (4), причем не обязательно различным. Например, $d_2=5$ попадает в класс A_5 , $d_5 = h_7=7$ попадают в класс A_7 . Перейдем от операций над элементами (3) классов к операциям над самими классами (4) как над некоторыми множествами.

Эти операции введем следующим образом.

1. Сложение:

$$A_k + A_l = \begin{cases} A_{k+l}, & k+l < n \\ A_{k+l-n}, & k+l \geq n \end{cases} \quad (5)$$

2. Вычитание:

$$A_k - A_l = \begin{cases} A_{k-l}, & k > l \\ A_{n+k-l}, & k \leq l \end{cases} \quad (6)$$

3. Умножение:

$$A_k A_l = A_b, \quad kl - mn = b, \quad 0 \leq b < n - 1 \quad (7)$$

Введенные операции сложения и умножения обладают коммутативностью и ассоциативностью, кроме того, умножение дистрибутивно по отношению к сложению.

С учетом введенных операций (5), (6), (7) для классов (4) аналогично (3) можно записать:

$$\sum_{i=1}^S A_{d_i} A_{h_i} = A_b \quad (8)$$

Например, для кода 34 2874 и весовых коэффициентов 1, 2, 3, 4, 5, 6 соотношение (8) можно записать в виде:

$$A_3A_1 + A_4A_2 + A_2A_3 + A_8A_4 + A_7A_5 + A_4A_6 = A_9.$$

Следовательно, указанный код в (4) относится к классу A_9 , что согласуется со значением остатка $b=9$, вычисленным для кода 34 2874 по формуле (3).

Для выявления ошибки в коде в соотношении (8) должен измениться остаток b , т.е. сумма в (8) должна перейти в (4) из одного класса в другой.

Наиболее распространенной ошибкой, связанной с человеческим фактором, является изменение цифры в одном разряде кода.

Например, вместо правильного кода 29 5316 записано 29 5816. В результате такой ошибки одно из слагаемых в выражении (8) должно перейти из некоторого класса A_l в класс A_k , и, следовательно, вся сумма выражения (8) перейдет в другой класс. Пусть в этом слагаемом весовой коэффициент $h_i \in A_k$, неискажённая цифра кода $d_i \in A_l$, искажённая цифра кода $d'_i \in A_l$.

Тогда не должно выполняться следующее равенство

$$A_k A_l = A_k A_l,$$

или

$$A_k (A_l - A_l') = A_0. \quad (9)$$

Равенство (9) нарушается, если каждый сомножитель в левой части (9), а также их произведение не равны нулю (A_0).

Чтобы $A_k \neq A_0$, любой весовой коэффициент не должен быть равен нулю (или кратен модулю).

Произведение в левой части равенства (9) для любых h_i и d_i не будет равно нулю, если модуль n простое число. Это следует из того, что произведение двух целых чисел, меньших некоторого числа, не может делиться на это число без остатка, если оно простое. Отсюда следует, что модуль n должен быть простым числом.

Второй сомножитель в равенстве (9) не будет равен нулю, если $A_1 \neq A_1$, т.е. искаженные и неискаженные цифры разрядов кода всегда окажутся в разных классах. При цифровых десятичных кодах разряды в коде могут принимать любое значение от 0 до 9 включительно, т.е. в (4) должно быть не менее 10 различных классов, а, следовательно, модуль должен быть больше 10. Модуль не может быть равным 10, так как 10 – число не простое. Ближайшим простым числом, большим 10, является 11, которое и должно быть принято в качестве модуля. Использование в качестве модуля следующих после 11 простых чисел, например 13, при однозначном контрольном числе приводит к нежелательному появлению новых двузначных остатков 11 и 12.

Отметим также, что при значении модуля меньше 10 (например, $n = 7$) и при искажении в разряде кода, заключающемся в замене цифры 2 на цифру 9, ошибка не будет обнаружена, так как обе цифры относятся к одному и тому же классу A_2 (по модулю $n = 7$). Например, при использовании модуля $n = 7$ и весовых коэффициентов 1, 2, 3, 4, 5, 6 для правильного кода 56 2321 и искаженного кода 56 2391 остаток, вычисленный по (3), получается равным 2 и ошибка не обнаруживается.

Таким образом, для защиты от ошибок рассмотренного типа необходимо выполнение следующие условия: модуль должен быть равен 11, весовые коэффициенты не должны быть равны нулю (или кратны модулю).

Для нахождения дополнительных требований к весовым коэффициентам предположим, что произошла ошибка, заключающаяся в перестановке

цифр, находящихся на i -м и j -м разрядах кода. Например, вместо правильного кода $\underline{39}\underline{45}16$ записано $\underline{35}\underline{49}16$.

В этом случае для выявления ошибок необходимо, чтобы сумма произведений весовых коэффициентов и цифр на i -м и j -м разрядах кода после искажения переходила в другой класс в (4).

Пусть на i -м и j -м разрядах кода весовые коэффициенты $h_i \in A_{k_i}$ и $h_j \in A_{k_j}$ цифры разрядов кода $d_i \in A_{l_i}$ и $d_j \in A_{l_j}$, тогда не должно выполняться равенство

$$A_{k_i} A_{l_i} + A_{k_j} A_{l_j} = A_{k_i} A_{l_j} + A_{k_j} A_{l_i}$$

или

$$(A_{k_i} - A_{k_j})(A_{l_i} - A_{l_j}) = A_0. \quad (10)$$

В общем случае $A_{l_i} \neq A_{l_j}$, поэтому чтобы равенство (10) не выполнялось, требуется в дополнение к условиям, перечисленным выше, принять ещё одно условие

$$A_{k_i} \neq A_{k_j} \quad (11)$$

Условие (11) означает, что весовые коэффициенты должны принадлежать к различным классам в (4). На основании этого в качестве весовых коэффициентов может быть принят любой набор чисел, принадлежащих различным классам в (4). Преимуществом данного результата является возможность его использования для практического построения набора весовых коэффициентов.

При этом следует учитывать, что каждый из классов в (4) содержит много чисел, равнозначных с учетом выражения (8) при их использовании в качестве весовых коэффициентов. Например, для модуля $n = 11$ одинаковый результат дает выбор в качестве весового коэффициента 100 или 1, поскольку они оба относятся к классу A_1 (т.е. дают при делении на 11 в остатке 1).

С этих позиций рассмотрим код как некоторое десятичное число и представим его в виде суммы согласно равенству (8).

Например,

$$4832 = 10^3 \cdot 4 + 10^2 \cdot 8 + 10^1 \cdot 3 + 1 \cdot 2.$$

В этом равенстве весовые коэффициенты равны

$$h_1 = 10^3, h_2 = 10^2, h_3 = 10, h_4 = 1$$

Для модуля $n = 11$: $h_1 = 10^3$ в остатке дает 10, $h_2 = 10^2$ в остатке дает 1, $h_3 = 10$ в остатке дает 10, $h_4 = 1$ в остатке дает 1.

Это означает, что непосредственное деление на модуль десятичного числа, соответствующего коду, эквивалентно построению суммы согласно равенству (8) с набором весовых коэффициентов 10, 1, 10, 1, ..., которые, однако, не удовлетворяют условиям (11).

Например, при записи вместо правильного кода 4832 искаженного кода 5822 и их непосредственном делении на модуль $n = 11$ остаток получается одинаковым ($b = 3$), т.е. ошибка не обнаруживается. Поэтому для вычисления контрольного числа и предусмотрен специальный алгоритм, реализованный выражениями (1) и (2).

Если обозначить максимально допустимую длину кода через T , то из условия (11) принадлежности коэффициентов для всех разрядов различным классам (4) следует, что должно выполняться неравенство:

$$T < n-1 \tag{12}$$

Условие (12) означает, что при модуле $n = 11$ и однозначном контрольном числе максимальная длина кода, защищаемого от ошибок, может быть не более 10.

Кроме того, важно обратить внимание на то, что при выборе модуля $n = 11$ двузначное контрольное число получается только в одном случае, когда $b = 10$. Поэтому для обеспечения однозначного контрольного числа необходимо исключить этот случай. В частности, в принятой для общероссийских

классификаторов методике расчета контрольного числа [1] рекомендуется проводить повторный расчет контрольного числа с последовательностью весовых коэффициентов, сдвинутой на два разряда влево.

Рассмотрим, для каких кодов это не дает желаемого результата.

Равенство (8) применительно к ОКП, в котором длина кода равна 6, можно записать в виде:

$$A_1 A_{d_1} + A_2 A_{d_2} + \dots + A_6 A_{d_6} = A_b \quad (13)$$

После сдвига последовательности весовых коэффициентов на два разряда влево получаем:

$$A_3 A_{d_1} + A_4 A_{d_2} + \dots + A_8 A_{d_6} = A_{b'} \quad (14)$$

Контрольные числа будут одинаковы, если правые части в равенствах (13) и (14) будут равны, т.е. $A_b = A_{b'}$. Вычитая из равенства (14) равенство (13), получаем:

$$(A_3 - A_1) A_{d_1} + (A_4 - A_2) A_{d_2} + \dots + (A_8 - A_6) A_{d_6} = A_0$$

с учетом операции (6) имеем

$$A_2 (A_{d_1} + A_{d_2} + \dots + A_{d_6}) = A_0$$

или

$$A_{d_1} + A_{d_2} + \dots + A_{d_6} = A_0 \quad (15)$$

Условие (15) означает, что при применении основной и сдвинутой последовательности весовых коэффициентов контрольное число не изменяется, если сумма цифр разрядов кода равна или кратна модулю.

Например, для кода 19 6240, сумма цифр которого делится на 11, контрольное число равно 10 как для основной, так и для сдвинутой на два разряда влево последовательности весов. Условие (15) следует использовать при необходимости исключения кодов, для которых получается двузначное контрольное число, равное 10.

Заменяя (14) на

$$A_2 A_{d_1} + A_3 A_{d_2} + \dots + A_7 A_{d_6} = A_b''$$

или на

$$A_4 A_{d_1} + A_5 A_{d_2} + \dots + A_9 A_{d_6} = A_b'''$$

и проводя далее аналогичные действия, можно показать, что контрольные числа будут одинаковыми также при использовании последовательностей весовых коэффициентов, сдвинутых не только на два разряда влево, но и на один или на три разряда влево в случае соответствия кода условию (15).

Проведенный анализ, основанный на построении классов чисел и операций над ними, позволил математически обосновать требования к выбору модуля и весовых коэффициентов, а также определить условие, при котором значение контрольного числа не изменяется при сдвиге последовательности весовых коэффициентов.

Список использованной литературы

1. Основные положения и порядок проведения работ по разработке, ведению и применению общероссийских классификаторов // Правила стандартизации ПР 50.1.024-2005. – М.: СТАНДАРТИНФОРМ, 2006. – 34 с.