УДК 621.327

МЕТОД ДВУХПРИЗНАКОВОГО БИНОМИНАЛЬНОГО КОДИРОВАНИЯ ДВОИЧНЫХ ДАННЫХ

к.т.н. В.В. Баранник
(представил проф. А.В. Королёв)

Излагается метод компактного представления двоичных данных на основе двумерного (двухпризнакового) биноминального кодирования. В этом случае сжатие изображений достигается за счет устранения структурной избыточности, обусловленной ограниченным числом единичных элементов в двоичных сериях.

Введение. Большие объемы двоичных данных, обрабатываемых в информационно-вычислительных системах, приводят к снижению скорости обработки, к росту количества вычислительных ресурсов и повышению времени доведения информации по каналам связи. Одним из направлений выхода из перечисленных трудностей является устранение избыточности в двоичных информационных потоках по их структурным признакам [1]. Для этого требуется разработать метод компактного представления двоичных данных.

Постановка задачи. Необходимо разработать метод сжатия на основе представления двоичных данных биноминальными кодами. При этом требуется обеспечить исключение двумерной структурной избыточности с нулевой погрешностью.

Обобщенная схема двухпризнакового биноминального кодирования двоичных данных. Биноминальным двоичным кодированием (БДК) является такое кодирование двоичных последовательностей, в результате которого формируется число, верхней границей которого является биноминальный коэффициент. В общем случае биноминальное кодирование по одному признаку задается формулой

$$N(\chi)_j = \sum_{i=1}^{m} a_{ij} V(\chi)_i \ , \tag{1}$$

где $N(\chi)_j$ – биноминальное число (БЧ) [1, 2], полученное для j-й двоичной последовательности, длиной m элементов; a_{ij} – ij-й двоичный элемент; $V(\chi)_i$ – весовой коэффициент a_{ij}-го элемента, зависящий от признака χ [1, 2].

В силу определения ДБК выполняется неравенство

$$N(\chi)_j \le V(m, \chi_j) - 1, \qquad (2)$$

где $V(m, \chi_j)$ – количество двоичных последовательностей, для которых значение признака χ равно χ_j.

В качестве признаков выбираются структурные характеристики, что обусловлено возможностью их количественной оценки. Причем для биноминального кодирования признаком χ должен быть такой структурный признак, значения которого делят все двоичное множество $\Omega(2,m)$ на k непересекаемых классов. Тогда выполняется равенство

$$\Omega(2,m) = \bigcup_{\chi_j=\chi_{\min}}^{\chi_{\max}} \Psi(m, \chi_j), \qquad (3)$$

где $\Psi(m, \chi_j)$ – множество двоичных последовательностей длиной m элементов со значением структурного признака, равным χ_j.

Мощность множества $\Psi(m, \chi_j)$ равна $V(m, \chi_j)$; $\chi_{\max}$, $\chi_{\min}$ – соответственно максимальное и минимальное значения признака χ, причем $\chi_{\max} - \chi_{\min} = k$.

Дополнительное увеличение коэффициента сжатия данных достигается при переходе от однопризнакового биноминального кодирования к двухпризнаковому. По аналогии с однопризнаковым для двухпризнакового биноминального кодирования выполняется неравенство

$$N(\chi_1, \chi_2)_j \le V(m, \chi_{1,j}, \chi_{2,j}), \qquad (4)$$

а выражение для формирования двухпризнакового БЧ $N(\chi_1, \chi_2)_j$ двоичных данных принимает вид

$$N(\chi_1, \chi_2)_j = \sum_{i=1}^{m} a_{ij} V(\chi_1, \chi_2)_i, \qquad (5)$$

где $N(\chi_1, \chi_2)_j$ – двухпризнаковое биноминальное число, полученное для j-й двоичной последовательности, длиной m элементов; $V(\chi_1, \chi_2)_i$ – весовой коэффициент a_{ij}-го элемента, зависящий от значений признаков χ_1 и χ_2 [1, 2]; $V(m, \chi_1, \chi_2)_i$ – количество двоичных последовательностей, для которых значения структурных признаков χ_1 и χ_2 соответственно равны $\chi_{1,j}$ и $\chi_{2,j}$.

В этом случае значения структурных признаков, равные $\chi_{1,j}$ и $\chi_{2,j}$, порождают двоичное множество $\Psi(m, \chi_{1,j}, \chi_{2,j})$. При этом для обеспечения взаимнооднозначного кодирования необходимо, чтобы выполнялось соотношение

$$\Omega(2,m) = \bigcup_{\chi_{1,j}=\chi_{1,\min}}^{\chi_{1,\max}} \bigcup_{\chi_{2,j}=\chi_{2,\min}}^{\chi_{2,\max}} \Psi(m, \chi_{1,j}, \chi_{2,j}), \qquad (6)$$

где множество $\Psi(m, \chi_{1,j}, \chi_{2,j})$ формируется как пересечение двух множеств

$$\Psi(m, \chi_{1,j}) \bigcap \Psi(m, \chi_{2,j}) = \Psi(m, \chi_{1,j}, \chi_{2,j}); \qquad (7)$$

$$\chi_{1,j} = \overline{\chi_{1,\min}, \chi_{1,\max}}; \quad \chi_{2,j} = \overline{\chi_{2,\min}, \chi_{2,\max}};$$

$\Psi(m, \chi_{1,j}), \Psi(m, \chi_{2,j})$ – множества двоичных последовательностей со значениями признаков, соответственно равными $\chi_{1,j}$ и $\chi_{2,j}$.

Выражения для нахождения весовых коэффициентов $V(\chi_1, \chi_2)_i$, зависят от выбора конкретных структурных признаков. Для того, чтобы при переходе от однопризнакового кодирования к двухпризнаковому кодированию обеспечивалось дополнительное повышение степени компактного представления, должны выполняться условия:

$$V(m, \chi_{1,j}, \chi_{2,j}) \le V(m, \chi_{1,j}); \qquad (8)$$

$$V(m, \chi_{1,j}, \chi_{2,j}) \le V(m, \chi_{2,j}), \qquad (9)$$

т.е. мощность множества $\Psi(m, \chi_{1,j}, \chi_{2,j})$ не должна превышать мощностей множеств $\Psi(m, \chi_{1,j}), \Psi(m, \chi_{2,j})$.

Значит для организации взаимооднозначного двухпризнакового биноминального кодирования и обеспечения дополнительного повышения степени компактного представления данных в качестве структурных признаков требуется выбирать такие признаки, которые будут удовлетворять условиям (6) – (9).

Разработка двухпризнакового биноминального кодирования двоичных данных. Выберем в качестве структурных признаков число ϑ двоичных серий и количество χ единичных элементов в двоичных последовательностях. В этом случае после выполнения математических преобразований следует, что выбранные признаки удовлетворяют условиям (6) – (9), а выражение для формирования двухпризнаковых биноминальных чисел задается формулой

$$N(\vartheta, \chi)_j = \sum_{i=1}^{m} a_{ij} V(\vartheta, \chi)_i \ . \qquad (10)$$

При этом значения весовых коэффициентов $V(\vartheta, \chi)_i$ находятся по формуле

$$V(\vartheta, \chi)_i = \binom{\chi_{i-1,j} - 1}{\vartheta_{i-1,j} - 1}\binom{m - i - \chi_{i-1,j} + 1}{\vartheta_{i-1,j}}, \qquad (11)$$

где $\chi_{i-1,j}$, $\vartheta_{i-1,j}$ – соответственно значения количества единиц и числа серий на $(i-1)$-м шаге кодирования j-й двоичной последовательности.

Значения величин $\chi_{i-1,j}$ и $\vartheta_{i-1,j}$ находятся рекурсивно по правилу:

– если $a_{ij} = 0$, то $\chi_{i,j} = \chi_{i-1,j}$ и $\vartheta_{i,j} = \vartheta_{i-1,j}$;

– если $a_{ij} = a_{i+1,j} = 1$, то $\chi_{i,j} = \chi_{i-1,j} - 1$ и $\vartheta_{i,j} = \vartheta_{i-1,j}$;

– если $a_{ij} = 1$, а $a_{i+1,j} = 0$, то $\chi_{i,j} = \chi_{i-1,j} - 1$ и $\vartheta_{i,j} = \vartheta_{i-1,j} - 1$.

Начальные значения признаков соответственно равны $\chi_{0,j} = \chi$ и $\vartheta_{0,j} = \vartheta$.

Таким образом, выражения (10), (11) и правило рекурсивного нахождения значений признаков организуют двухпризнаковое биноминальное кодирование двоичных данных по количеству единиц в двоичных сериях.

Рассмотрим случай, когда количество единиц χ в двоичных сериях изменяется в пределах $\chi_\ell \le \chi \le \chi_h$, где χ_ℓ, χ_h – соответственно нижний и верхний пределы изменения значения χ. Тогда выражения для определения двухпризнакового БЧ $N(\vartheta, X)_j$ будет иметь вид

$$N(\vartheta, X)_j = \sum_{\chi=\chi_\ell}^{\chi_h} \sum_{i=1}^{m} a_{ij} V(\vartheta, \chi)_i \ , \qquad (12)$$

где X – вектор значений признака χ:

$$X = \{\chi_\ell, \dots, \chi_j, \dots, \chi_h\}.$$

Если значения признака ϑ также изменятся в некоторых пределах $\vartheta_\ell \le \vartheta \le \vartheta_h$, где ϑ_ℓ, ϑ_h – соответственно нижний и верхний пределы изменения значения ϑ, то величина двухпризнакового БЧ $N(\Theta, X)_j$ вычисляется на основе выражения

$$N(\Theta, X)_j = \sum_{\vartheta=\vartheta_\ell}^{\vartheta_h} \sum_{\chi=\chi_\ell}^{\chi_h} \sum_{i=1}^{m} a_{ij} V(\vartheta, \chi)_i \ , \tag{13}$$

где Θ – вектор значений признака ϑ:

$$\Theta = \{\vartheta_\ell, ..., \vartheta_j, ..., \vartheta_h\} .$$

Таким образом, выражение (13) позволяет определить значение двухпризнакового биноминального числа для двоичной последовательности тогда, когда значения признаков изменяются в некоторых пределах.

Сравнительная оценка по степени сжатия двоичных данных. В качестве сравниваемых методов использовались разработанный метод и методы арифметического кодирования двоичных данных [3]. В качестве тестов использовались фрагменты фотографических изображений на разрядном уровне. Проведенные эксперименты показали, что степень сжатия двоичных данных для разработанного метода больше, чем степень сжатия для известных методов в среднем в **2** раза.

Выводы. 1. Разработан метод двухпризнакового биноминального кодирования двоичных данных по количеству единиц в двоичных сериях. В этом случае сжатие данных обеспечивается за счет устранения структурной избыточности, обусловленной ограниченным количеством единичных элементов в двоичных сериях.

2. Разработано двухпризнаковое биноминальное кодирование двоичных данных на случай, когда значения структурных признаков изменяются в некоторых пределах.

3. Эксперименты с массивами видеоданных показали, что выигрыш по степени компактного представления двоичных данных для разработанного метода кодирования относительно степени компактного представления для известных методов в среднем равен **2** разам.

ЛИТЕРАТУРА

1. *Королёв А.В., Баранник В.В. Оценка количества информации изображения по числу серий одинаковых элементов // Системи обробки інформації. – Х.: НАНУ, ПАНМ, ХВУ. – 2002. – Вип. 2(18). – С. 43 – 46.*
2. *Корн Г., Корн Т. Справочник по математике. – М.: Наука, 1973. – 831 с.*
3. *Зубарев Ю.В., Дворкович В.П. Цифровая обработка телевизионных и компьютерных изображений – М.: МЦНТИ, 1997. – 212 с.*

Поступила 13.09.2002

Баранник Владимир Викторович, *канд. техн. наук, старший научный сотрудник информационно-вычислительного центра ХВУ. В 1994 году окончил ХВУ. Область научных интересов – обработка и передача информации.*