KYBERNETIKA ČÍSLO 5, ROČNÍK 4/1968

К проблеме однозначности кодирования

Иван Гавел

Известно, что проблема однозначности кодовой системы алгорифмически разрешима. Ниже доказывается, что эта-же проблема является алгорифмически неразрешимой для кодовых систем высшего порядка, т. е. таких, которые сопоставляют исходному — кодированному — слову k-членную систему слов ($k \geqslant 2$).

Кодовая система $\mathscr{K} = \langle \mathsf{A}, \mathsf{B}, \varphi \rangle$ задана своими конечными алфавитами A, B и отображением φ множества A в $\mathsf{B}^* - \{\Lambda\}$ (Λ обозначает пустое слово). Для любого непустого слова P в алфавите A, $P = \xi_1 \dots \xi_n$, кодом P называют следующее слово в алфавите B

$$\Phi(P) = \varphi(\xi_1)\, \varphi(\xi_2) \dots \varphi(\xi_n)\,.$$

Кодовая система $\mathscr{K}$ однозначна, если для любых слов P, Q в A имеет место

$$P \neq Q \Rightarrow \Phi(P) \neq \Phi(Q)\,.$$

Эти определения хорошо известны (см. [2], [4]); также известно, что существует единый общий метод (алгорифм), позволяющий для любой кодовой системы решить, однозначна-ли она или нет (см. [4], [5]).

Дадим немного более общее определение кодовой системы: кодовая система $\mathscr{K}$ порядка k ($k \geqslant 1$) задана посредством $(2k + 1)$-членной последовательности

$$\langle \mathsf{A}, \mathsf{B}_1, \dots, \mathsf{B}_k, \varphi_1, \dots, \varphi_k \rangle$$

где A, B_i ($1 \leqq i \leqq k$) конечные алфавиты и φ_i отображения множества A в B_1^*. Для любого $\xi \in \mathsf{A}$ приэтом существует по крайней мере одно i такое, что $\varphi_i(\xi) \neq \Lambda$.

Кодом $\Phi(P)$ слова $P = \xi_1 \dots \xi_n$ в A называют упорядоченную k-членную последовательность

$$\langle \varphi_1(\xi_1) \dots \varphi_1(\xi_n), \varphi_2(\xi_1) \dots \varphi_2(\xi_n), \dots, \varphi_k(\xi_1) \dots \varphi_k(\xi_n) \rangle\,.$$

Система $\mathscr{K}$ однозначна, если для любых слов P, Q в A имеет место

$$P \neq Q \Rightarrow \Phi(P) \neq \Phi(Q)\,.$$

Проблема однозначности кодовых систем порядка $k > 1$ в отличие от „одномерного" случая алгорифмически неразрешима. Имеет место следующая

Теорема. *Невозможен алгорифм, который для любой кодовой системы порядка* k ($k > 1$) *решает, однозначна-ли она или нет.*

Доказательство вытекает из нескольких лемм, которые мы приводим ниже.

Лемма 1. *Пусть* $k > 2$. *Если проблема однозначности алгорифмически разрешима для кодовых систем порядка* k, *то она также алгорифмически разрешима для кодовых систем порядка* 2.

Доказательство вытекает из следующего факта: для любой кодовой системы $\mathscr{K}$ порядка 2 может быть построена кодовая система $\mathscr{K}'$ порядка k, которая однозначна в том и только в том случае, когда $\mathscr{K}$ является однозначной. Действительно, если $\mathscr{K} = \langle \mathsf{A}, \mathsf{B}_1, \mathsf{B}_2, \varphi_1, \varphi_2 \rangle$, положим

$$\mathscr{K}' = \langle \mathsf{A}, \mathsf{B}_1, \mathsf{B}_2, \mathsf{B}_1, \ldots, \mathsf{B}_1, \varphi_1, \varphi_2, \varphi_1, \ldots, \varphi_1 \rangle\,.$$

$\mathscr{K}'$ обладает требуемым свойством.

Из леммы 1 вытекает, что для доказательства теоремы достаточно доказать невозможность алгорифмического решения проблемы однозначности для кодовых систем порядка 2. Это и является содержанием следующей леммы.

Лемма 2. *Невозможен алгорифм, который для любой кодовой системы порядка* 2 *решает, однозначна-ли она или нет.*

Доказательство проведем сведением к общей комбинаторной проблеме Поста. Покажем, что существование такого алгорифма означало-бы разрешимость общей проблемы Поста, что однако, невозможно.

Общая комбинаторная проблема Поста:

для любой системы пар слов в алфавите C

(1) $$\Pi = \{(c_i, d_i);\ i = 1, \ldots, m\}$$

решить, существует-ли последовательность натуральных чисел

(2) $$I = \{i_1, \ldots, i_n\} \quad (1 \leqq i_j \leqq m)$$

такая, что

(3) $$c_{i_1} c_{i_2} \ldots c_{i_n} = d_{i_1} d_{i_2} \ldots d_{i_n}\,.$$

Известно ([6], из последних работ напр. [1], [3]), что общая комбинаторная проблема Поста является алгорифмически неразрешимой.

Напомним, что если для системы (1) существует последовательность (2) удовлетворяющая (3), то говорят, что проблема Поста (единичная) для (1) разрешима ((2) является решением для (1) и таких решений существует бесконечно много), в обратном случае говорят, что проблема Поста для (1) неразрешима.

Лемма 3. *По любой системе пар слов*

$$\Pi = \{(c_i, d_i);\ i = 1, \dots, m\}$$

в алфавите C *может быть построена кодовая система* $\mathscr{K}$ *порядка* 2, *которая является однозначной в том и только в том случае, когда проблема Поста* (*единичная*) *для* Π *неразрешима.*

Доказательство. Положим

$$\mathscr{K} = \langle \mathsf{A}, \mathsf{B}_1, \mathsf{B}_2, \varphi_1, \varphi_2 \rangle$$

где

$$\mathsf{A} = \{a_1, \dots, a_m, A_1, \dots, A_m, b_1, \dots, b_m, B_1, \dots, B_m, A, B\},$$
$$\mathsf{B}_1 = \mathsf{C} \cup \{\vdash, \dashv\},\ \mathsf{B}_2 = \{e_1, \dots, e_m, \vdash, \dashv, *\}$$

и φ_i заданы следующим образом:

$$\begin{aligned}
\varphi_1(a_i) &= \vdash c_i, & \varphi_2(a_i) &= \vdash * e_i,\\
\varphi_1(\mathsf{A}_i) &= c_i, & \varphi_2(\mathsf{A}_i) &= * e_i,\\
\varphi_1(\mathsf{A}) &= \dashv, & \varphi_2(A) &= * \dashv,\\
\varphi_1(b_i) &= \vdash d_i, & \varphi_2(b_i) &= \vdash * e_i *,\\
\varphi_1(B_i) &= d_i, & \varphi_2(B_i) &= e_i *,\\
\varphi_1(B) &= \dashv, & \varphi_2(B) &= \dashv.
\end{aligned}$$

Пусть система натуральных чисел

$$I = \{i_1, \dots, i_n\} \quad (1 \leqq i_j \leqq m)$$

является решением для Π. Покажем, что $\mathscr{K}$ неоднозначна.

Имеем

$$\Phi(a_{i_1} A_{i_2} \dots A_{i_n} A) = \langle \vdash c_{i_1} c_{i_2} \dots c_{i_n} \dashv, \vdash * e_{i_1} * e_{i_2} * \dots * e_{i_n} * \dashv \rangle =$$
$$= \langle \vdash d_{i_1} d_{i_2} \dots d_{i_n} \dashv, \vdash * e_{i_1} * e_{i_2} * \dots * e_{i_n} * \dashv \rangle = \Phi(b_{i_1} B_{i_2} \dots B_{i_n} B).$$

Пусть наоборот $\mathscr{K}$ не является однозначной. Следовательно, существуют $\xi_1, \dots, \xi_{n'}, \eta_1, \dots, \eta_{n''} \in \mathsf{A}$ такие, что

(4) $$\Phi(\xi_1, \dots, \xi_{n'}) = \Phi(\eta_1 \dots \eta_{n''}), \quad \xi_1 \dots \xi_{n'} \neq \eta_1 \dots \eta_{n''}.$$

Мы можем считать, что выполнено следующее условие:

(4′) если $1 \leqq l' \leqq n'$, $1 \leqq l'' \leqq n''$ и $\langle l', l'' \rangle \neq \langle n', n'' \rangle$,

то

$$\Phi(\xi_1 \dots \xi_{l'}) \neq \Phi(\eta_1 \dots \eta_{l''}).$$

Из (4) вытекает существование i_1 такого, что либо

(5) $$\xi_1 = a_{i_1} \quad \text{и} \quad \eta_1 = b_{i_1}$$

либо

(6) $$\xi_1 = b_{i_1} \quad \text{и} \quad \eta_1 = a_{i_1}\,.$$

Без ограничения общности предположим, что имеет место (5). Докажем, что $n' = n''$ и что либо $n' = 2$ и $\xi_2 = A$, $\eta_2 = B$ либо $n' > 2$ и существуют $i_2, \ldots, \ldots, i_{n'-1}$ такие что

$$\begin{aligned} \xi_2 &= a_{i_2}\,, & \eta_2 &= b_{i_2}\,,\\ \xi_{n'-1} &= a_{i_{n'-1}}\,, & \eta_{n'-1} &= b_{i_{n'-1}}\,,\\ \xi_{n'} &= A\,, & \eta_{n'} &= B\,. \end{aligned}$$

Имеет место

$$\Phi(\xi_1) = \langle \vdash c_{i_1}, \vdash * e_{i_1} \rangle\,, \quad \Phi(\eta_1) = \langle \vdash d_{i_1}, \vdash * e_{i_1} * \rangle\,,$$

следовательно

$$\Phi(\xi_1) \neq \Phi(\eta_1) \quad \text{и} \quad n' \geqslant 2\,, \quad n'' \geqslant 2\,.$$

Из (4) вытекает, что либо $\xi_2 \in \{A_i\}$ либо $\xi_2 = A$, других возможностей нет.

а) пусть $\xi_2 = A$; тогда

$$\Phi(\xi_1\,\xi_2) = \langle \vdash c_{i_1} \dashv, \vdash * e_{i_1} * \dashv \rangle\,, \quad \eta_2 = B\,,$$

следовательно

$$\Phi(\xi_1\xi_2) = \Phi(\eta_1\eta_2)\,.$$

Учитывая предположение (4'), мы получаем $n' = n'' = 2$. На основании (4) заключаем

$$c_{i_1} = d_{i_1}$$

следовательно, для Π существует решение $I = \{i_1\}$.

б) допустим, что существует i_2 такое, что $\xi_2 = A_{i_2}$, следовательно $n' > 2$. Но тогда обязательно $\eta_2 = B_{i_2}$, $\Phi(\xi_1\xi_2) \neq \Phi(\eta_1\eta_2)$ и $n'' > 2$.

Докажем, что для любого u $(1 < u \leqq n' - 1)$ прежде всего $u < n''$ а далее, что существуют $i_1, \ldots, i_u$ такие что

$$\begin{aligned} \xi_1 &= a_{i_1}\,, & \eta_1 &= b_{i_1}\,,\\ \xi_2 &= A_{i_2}\,, & \eta_2 &= B_{i_2}\,,\\ &\ldots\ldots\ldots\ldots\ldots\ldots\\ \xi_u &= A_{i_u}\,, & \eta_u &= B_{i_u}\,. \end{aligned}$$

Это верно для $u = 2$. Допустим, что утверждение имеет место для некоторого $u < n' - 1$. Докажем, что оно имеет место тогда и для $u + 1$.

Имеем

$$\Phi(\xi_1 \dots \xi_u) = \langle \vdash c_{i_1} c_{i_2} \dots c_{i_u}, \quad \vdash * e_{i_1} * e_{i_2} * \dots * e_{i_u} \rangle,$$
$$\Phi(\eta_1 \dots \eta_u) = \langle \vdash d_{i_1} d_{i_2} \dots d_{i_u}, \quad \vdash * e_{i_1} * e_{i_2} * \dots * e_{i_u} * \rangle.$$

Очевидно, либо $\xi_{u+1} = A$ либо существует i_{u+1} такое что $\xi_{u+1} = A_{i_{u+1}}$. Первое, однако, невозможно, так как оно влечет $\eta_{u+1} = B$ и

$$\Phi(\xi_1 \dots \xi_{u+1}) = \Phi(\eta_1 \dots \eta_{u+1}),$$

что противоречит (4′). Следовательно, $\xi_{u+1} = A_{i_{u+1}}$ для некоторого i_{u+1}, но тогда $\eta_{u+1} = B_{i_{u+1}}$,

$$\Phi(\xi_1 \dots \xi_{u+1}) \neq \Phi(\eta_1 \dots \eta_{u+2}) \quad \text{и} \quad n'' > u + 1.$$

В общем мы заключаем, что $n' - 1 < n''$ и существуют $i_1, \dots, i_{n'-1}$ такие что

$$\begin{array}{ll} \xi_1 = a_{i_1}, & \eta_1 = b_{i_1}, \\ \xi_2 = A_{i_2}, & \eta_2 = B_{i_2}, \\ \dots\dots\dots\dots\dots & \\ \xi_{n'-1} = A_{i_{n'-1}}, & \eta_{n'-1} = B_{i_{n'-1}}. \end{array}$$

Имеет место

$$\Phi(\xi_1 \dots \xi_{n'-1}) = \langle \vdash c_{i_1} \dots c_{i_{n'-1}}, \vdash * e_{i_1} * \dots * e_{i_{n'-1}} \rangle,$$
$$\Phi(\eta_1 \dots \eta_{n'-1}) = \langle \vdash d_{i_1} \dots d_{i_{n'-1}}, \vdash * e_{i_1} * \dots * e_{i_{n'-1}} * \rangle;$$

из (4) следует

$$\xi_{n'} = A, \quad \eta_{n'} = B \quad \text{и} \quad n'' = n.$$

Далее имеем из (4)

$$c_{i_1} \dots c_{i_{n'}} = d_{i_1} \dots d_{i_{n'}}$$

и следовательно, найдено решение $I = \{i_1, \dots, i_{n'}\}$ для Π.

Тем самым доказана лемма 3 и лемма 2 и теорема в целом. Если бы существовал алгорифм решающий для любой кодовой системы порядка 2 является-ли она или нет однозначной, было бы возможным алгорифмически решать общую комбинаторную проблему Поста.

(Поступило 2 го января 1968 г.)

ЛИТЕРАТУРА

[1] Floyd, R. W.: New proofs of old theorems in logic and formal linguistics. Carnegie Institute of Technology, Pennsylvania, November 1966.

[2] Левенштейн, В. И.: О некоторых свойствах кодовых систем. ДАН СССР *140* (1961), т. 6.

[3] Мальцев, А. И.: Алгоритмы и рекурсивные функции. Наука, Москва 1965.

[4] Марков, Ал. А.: Об алфавитном кодировании. ДАН СССР *132* (1960), 3.

[5] Марков, Ал. А.: Об алфавитном кодировании. ДАН СССР *139* (1961), 3.

[6] Post, E. L.: A variant of a recursively unsolvable problem, Bull. Amer. Math. Soc. *52* (1946), 164—268.

K problému jednoznačnosti kódování

IVAN HAVEL

V práci se zavádí pojem kódovací soustavy řádu k ($k \geqq 1$). Taková soustava je zadána $(2k + 1)$-ticí

$$\mathscr{K} = \langle \mathsf{A}, \mathsf{B}_1, \ldots, \mathsf{B}_k, \varphi_1, \ldots, \varphi_k \rangle ,$$

kde A, B_i jsou konečné abecedy a φ_j jsou zobrazení A do B_i^*.

Kódem $\Phi(P)$ slova $P = \xi_1 \ldots \xi_n$ ($\xi_i \in \mathsf{A}$) se rozumí uspořádaná k-tice

$$\varphi_1(\xi_1)\varphi_1(\xi_2) \ldots \varphi_1(\xi_n), \varphi_2(\xi_1) \ldots \varphi_2(\xi_n), \ldots, \varphi_k(\xi_1) \ldots \varphi_k(\xi_n) .$$

Soustava $\mathscr{K}$ je jednoznačná, jestliže $P \neq Q \Rightarrow \Phi(P) \neq \Phi(Q)$. V případě jednoduchého kódování ($k = 1$) existuje algoritmus dovolující o každé kódovací soustavě rozhodnout, zda je či není jednoznačná (viz [4]). V práci je redukcí k Postovu problému ukázáno, že problém jednoznačnosti je algoritmicky neřešitelný pro kódovací soustavy vyšších řádů ($k \geqq 2$).

Ivan Havel, Matematický ústav ČSAV, Žitná 25, Praha 1.