

УДК 510.62

М. Ф. БОНДАРЕНКО, канд. техн. наук,
Ю. П. ШАБАНОВ-КУШНАРЕНКО, д-р техн. наук

О МАТЕМАТИЧЕСКОМ ОПИСАНИИ ЕСТЕСТВЕННОГО ЯЗЫКА

Разработка формального аппарата теории интеллекта [1] открывает новые перспективы для решения задачи математического моделирования естественного языка. Будем смотреть на язык как на детерминированный, дискретный и конечный объект. Основную задачу мы видим в том, чтобы формализовать понятие текста путем математического описания отношения, выделяющего подмножество всех текстов из некоторого множества всевозможных

строчек ограниченной длины, составленных из букв конечного алфавита. Это отношение назовем текстовым. Текстовое отношение должно быть описано как отношение с параметром, поскольку решение вопроса о том, является ли данная строчка текстом или нет, зависит также и от того, кто эту строчку воспринимает. Будем рассматривать текст как многоуровневую конструкцию, полагая, что из букв образуются морфемы, из морфем — словоформы, из словоформ — предложения, из предложений — абзацы и т. д. Отдельные части этой конструкции — буквы, морфемы, словоформы, предложения и т. д. будем называть фрагментами текста.

Каждый фрагмент связан с окружающим его текстом, подвергает его воздействию. Действие отброшенной части текста может быть заменено введением некоторого набора признаков, который будем интерпретировать как смысл фрагмента текста. Смысл — это набор промежуточных переменных, связывающих уравнения, описывающие фрагмент текста, с уравнениями, описывающими остальную часть текста. Набор значений этих переменных интерпретируем как значение смысла. Смысл будем рассматривать как переменный вектор $X = (x_1, x_2, \dots, x_m)$, где m — число компонентов вектора X . Переменные x_1, x_2, \dots, x_m назовем компонентами смысла. Значения этих переменных назовем значениями компонентов смысла. Пусть y_1, y_2, \dots, y_n — некоторый фрагмент текста, где y_1, y_2, \dots, y_n — буквы, стоящие на 1, 2, ..., n местах фрагмента. Набор $Y = (y_1, y_2, \dots, y_n)$ буквенных переменных y_1, y_2, \dots, y_n назовем переменным фрагментом. Набор значений этих переменных назовем индивидуальным фрагментом или просто фрагментом текста. Аналогичным образом введем переменный текст и индивидуальный текст (просто текст). Любой конкретный фрагмент (текст) будем рассматривать как значение некоторого переменного фрагмента (текста). Переменные y_1, y_2, \dots, y_n назовем компонентами переменного фрагмента (переменного текста), их значения назовем значениями компонентов фрагмента (текста).

Между смыслом X и соответствующим ему переменным фрагментом Y существует определенная связь $L(X, Y)$. Назовем ее фрагментным отношением. Если фрагментом служит текст, то фрагментное отношение превращается в упоминавшееся ранее текстовое отношение. В роли параметров текстового отношения могут выступать любые из компонентов x_1, x_2, \dots, x_m смысла X . Будем придерживаться принципа однозначности, согласно которому текст (фрагмент) всегда однозначно зависит от его смысла. Принцип однозначности можно интерпретировать как требование полноты набора признаков (компонентов смысла), с помощью которого осуществляется выбор текста (фрагмента). Набор признаков назовем полным, если он при любых значениях признаков обеспечивает однозначность выбора соответствующего ему текста (фрагмента). Любой набор признаков, удовлетворяющий требова-

нию полноты по отношению к текстам (фрагментам), может быть принят в качестве смысла текста (фрагмента). Под неполным смыслом текста будем понимать набор, составленный из некоторых (не всех) компонентов смысла этого текста. Смысл же текста в принятом ранее значении будем называть полным смыслом текста.

Фрагментное отношение $L(X, Y)$, у которого в роли переменной Y выступает часть слова (например, буква, морфема) или целое слово, будем называть морфологическим отношением. В силу принципа однозначности любое морфологическое отношение $L(X, Y)$ есть функция зависимости фрагмента Y от его смысла, поэтому будем морфологическое отношение называть еще морфологической функцией. Пусть y_1 — первая буква окончания, область ее изменения зададим уравнением

$$y_1^d \vee y_1^e \vee y_1^{\bar{e}} \vee y_1^n \vee y_1^o \vee y_1^v \vee y_1^{\bar{v}} \vee y_1^y \vee y_1^{\bar{y}} \vee y_1^{\bar{y}} \vee y_1^{\bar{y}} = 1. \quad (1)$$

Смысл первой буквы окончания y_1 представим в виде n -мерного вектора (x_1, x_2, \dots, x_n) с буквенными компонентами x_1, x_2, \dots, x_n . Введем морфологическое отношение $L(x_1, x_2, \dots, x_n, y_1) = 1$ (2).

Разложим [2] предикат L (назовем его морфологическим предикатом) по переменной x_1

$$L(x_1, x_2, \dots, x_n, y_1) = x_1^a L(a_1, x_2, \dots, x_n, y_1) \vee x_1^{a_2} L(a_2, x_2, \dots, x_n, y_1) \vee \dots \vee x_1^{a_k} L(a_k, x_2, \dots, x_n, y_1). \quad (3)$$

При разложении полагаем, что переменная x_1 пробегает значения a_1, a_2, \dots, a_k . Принимая поочередно для переменной x_1 значения a_1, a_2, \dots, a_k , получаем k частных случаев морфологического предиката:

$$L_i(x_2, \dots, x_n, y_1) = L(a_i, x_2, \dots, x_n, y_1). \quad (4)$$

Отношения $L_i(x_2, \dots, x_n, y_1) = 1$ (5) могут изучаться отдельно, после чего из них можно собрать по формуле (3) предикат L :

$$L(x_1, x_2, \dots, x_n, y_1) = x_1^{a_1} L^*(x_2, \dots, x_n, y_1) \vee x_1^{a_2} L_2(x_2, \dots, x_n, y_1) \vee \dots \vee x_1^{a_k} L_k(x_2, \dots, x_n, y_1). \quad (6)$$

Если же предикаты L_1, L_2, \dots, L_k окажутся слишком сложными для непосредственного математического описания, то к ним тоже можно применить теорему о разложении и т. д. до тех пор, пока не придем к достаточно простым предикатам.

Сузим задачу введением пяти грамматических признаков, значения которых зафиксируем: $\xi_1 = o$ — вид морфемы со значением окончание, $\xi_2 = n$ — номер буквы окончания со значением первый, $\xi_3 = c$ — тип словоизменения со значением склонение, $\xi_4 = c$ — тип склонения со значением субстантивный, $\xi_5 = 0$ — вид словоформы со значением основной. Вводим переменные

признаки x_1 — падеж, x_2 — род, x_3 — число, x_4 — признак одушевленности, x_5 — признак ударности основы, x_6 — требование простановки знака $\cdot\cdot$ над буквой \ddot{e} , x_7 — последняя буква основы, x_8 — вид основы [3]. Области изменения введенных переменных задаем уравнениями:

$$x_1^H \vee x_1^P \vee x_1^N \vee x_1^B \vee x_1^T \vee x_1^I = 1, (7); \quad x_2^M \vee x_2^K \vee x_2^C = 1, (8);$$

$$x_3^E \vee x_3^M = 1, (9); \quad x_4^O \vee x_4^H = 1, (10); \quad x_5^Y \vee x_5^O = 1, (11);$$

$$x_6^C \vee x_6^H = 1, (12); \quad x_7^A \vee x_7^B \vee x_7^F \vee x_7^H \vee x_7^E \vee x_7^K \vee x_7^S \vee x_7^N \vee$$

$$\vee x_7^I \vee x_7^K \vee x_7^L \vee x_7^M \vee x_7^H \vee x_7^O \vee x_7^P \vee x_7^D \vee x_7^C \vee$$

$$\vee x_7^T \vee x_7^Y \vee x_7^F \vee x_7^X \vee x_7^H \vee x_7^N \vee x_7^U \vee x_7^M \vee$$

$$\vee x_7^O \vee x_7^H = 1, (13); \quad x_8^T \vee x_8^M = 1, (14).$$

Дальнейшая задача состоит в том, чтобы описать морфологическое отношение $L(x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, y) = 1$ (15). В силу полноты набора $(x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8)$ уравнение (15) задает некоторую функцию

$$y_1 = F(x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8). \quad (16)$$

Она будет описана, если мы представим ее в явном виде [4]:

$$y_1^\sigma = F_\sigma(x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8), \quad (17)$$

где $\sigma \in \{a, e, \ddot{e}, n, o, y, \text{ы}, \text{ю}, \text{я}, _ \}$.

В результате описания получаем следующие формулы:

$$y_1^a = (x_8^T s_1 \vee s_2 \vee s_3 \vee x_7^H) (x_3^E (x_1^N x_2^K \vee x_1^P (x_2^M \vee x_2^C) \vee x_2^M t_1) \vee \\ \vee x_3^M (t_2 x_2^C \vee x_1^H \vee x_1^T \vee x_1^I)) \quad (18); \quad y_1^e = x_6^C x_5^O (x_8^M s_1 x_3^E t_3 \vee \\ \vee (s_4 \vee x_7^H) (x_3^E t_3 \vee x_3^H (x_3^M \vee x_2^C) t_1)); \quad (19)$$

$$y_1^H = (x_8^M s_1 \vee s_2 \vee s_3 \vee s_4) (x_3^E x_1^P x_2^K \vee x_3^M t_2 (x_2^M \vee x_2^K)) \vee \\ \vee x_7^H (x_3^O (x_1^I x_5^Y \vee (x_1^P \vee x_1^I) x_2^K) \vee x_3^M t_2 (x_2^M \vee x_2^K)); \quad (20)$$

$$y_1^O = (x_8 s_1 \vee s_2) (x_3^E t_3 \vee x_3^M t_1 (x_2^M \vee x_5^O x_2^C)) \vee x_5^O x_1^Y \times \\ \times (x_3^E t_3 \vee x_3^M t_1 (x_2^M \vee x_2^C)) \vee x_5^O s_3 x_3^E t_4; \quad (21)$$

$$y_1^Y = (x_8^T s_1 \vee s_2 \vee s_3 \vee x_7^Y) x_3^E (x_1^H (x_2^M \vee x_2^C) \vee x_1^B x_2^K); \quad (22)$$

$$y_1^I = (x_8^T s_1 \vee x_7^Y) (x_3^E x_1^P x_2^K \vee x_3^M t_2 (x_2^M \vee x_2^K)); \quad (23)$$

$$y_1^O = (x_8^M s_1 \vee s_4 \vee x_7^Y) x_3^E (x_1^H (x_2^M \vee x_2^C) \vee x_7^B x_2^K); \quad (24)$$

$$\bar{y}_1 = x_5^Y (x_3^E t_2 x_2^M \vee x_3^M t_1 (x_2^K \vee x_2^C)). \quad (25)$$

В формулах (18) — (25) приняты следующие сокращения:

$$s_1 = x_7^O \vee x_7^H \vee x_7^L \vee x_7^M \vee x_7^H \vee x_7^M \vee x_7^H \vee x_7^P \vee x_7^D \vee x_7^C \vee x_7^T \vee x_7^F; \quad (26)$$

$$s_2 = x_7^r \vee x_7^k \vee x_7^x; \quad (27); \quad s_3 = x_7^{ж} \vee x_7^u \vee x_7^m \vee x_7^п; \quad (28)$$

$$s_4 = x_7^n \vee x_7^o \vee x_7^h \vee x_7^o \vee x_7^y \vee x_7^m \vee x_7^ю \vee x_7^a; \quad (29)$$

$$t_1 = x_1^p \vee x_1^b \vee x_1^o \quad (30); \quad t_2 = x_1^n \vee x_1^b \vee x_1^a \quad (31); \quad t_3 = x_1^r \vee (x_1^n \vee x_1^b) \vee x_2^c \quad (32).$$

Построим также модель для второй y_2 и третьей y_3 букв окончания основных словоформ субстантивного склонения. Имеем:

$$y_2^b \vee y_2^h \vee y_2^m \vee y_2^x \vee y_2^r = 1 \quad (33); \quad y_3^h \vee y_3^ю \vee y_3^r = 1 \quad (34).$$

Вводим дополнительный признак x_9 — вариант словоформы с областью [3]: $x_9^c \vee x_9^h = 1$ (35). Выражения для узнавания букв окончания имеют вид:

$$y_2^b = (x_8^m s_1 \vee s_2 \vee s_4 \vee x_7^u \vee x_7^h) x_3^m t_1 (x_2^m \vee x_5^o x_2^c); \quad (36)$$

$$x_2^h = x_1^r x_2^{ж} x_3^e x_9^h \vee (x_3^m s_1 \vee s_3) x_3^m t_1 (x_5^o \vee x_5^m x_5^a) \vee x_5^a (s_4 \vee x_7^h) (x_3^c t_2 x_2^m \vee x_3^m t_1 (x_2^{ж} \vee x_2^c)); \quad (37)$$

$$y_2^m = x_3^c x_1^r (x_2^m \vee x_2^c) \vee x_3^m (x_1^d \vee x_1^r) \quad (38); \quad y_2^x = x_3^m x_1^п \quad (39);$$

$$x_2^r = x_3^c ((x_1^h \vee x_1^b) (x_2^{ж} \vee x_2^c) \vee x_1^p \vee x_1^d \vee x_1^п \vee t_1 x_2^m \vee x_1^r x_2^{ж} x_9^c) \vee x_3^m t_2 \vee x_5^a (t_1 \vee t_3 \vee x_7^u) (x_3^c t_2 x_2^m \vee x_3^m t_1 (x_2^{ж} \vee x_2^c)); \quad (40)$$

$$x_3^h = x_1^r x_3^m \quad (41); \quad x_3^ю = x_1^r x_2^{ж} x_3^e x_9^c \quad (42);$$

$$x_3^r = x_1^h \vee x_1^p \vee x_1^d \vee x_1^п \vee x_1^r (x_2^m \vee x_2^c \vee x_2^{ж} x_9^h) x_3^e. \quad (43)$$

Список литературы: 1. Шабанов-Кушнарченко Ю. П. О теории интеллекта.— Проблемы бионики, 1979, вып. 22, с. 3—11. 2. Шабанов-Кушнарченко Ю. П. Об алгебре конечных предикатов.— АСУ и приборы автоматки, 1979, вып. 50, с. 14—20. 3. Математическое описание процесса склонения имен прилагательных/Ю. П. Шабанов-Кушнарченко, М. Ф. Бондаренко, В. М. Бондарев и др.— Проблемы бионики, 1980, вып. 24, с. 22—27. 4. Шабанов-Кушнарченко Ю. П. Аналитический метод явного задания конечных алфавитных операторов.— Проблемы бионики, 1980, вып. 25, с. 3—11.

Поступила 25 марта 1980 г.