УДК 519.7

Л.Л. КОЗЯЕВ, ХНУРЭ (г. Харьков)

ОБ АЛГЕБРОЛОГИЧЕСКОМ АППАРАТЕ МОРФОЛОГИЧЕСКОГО ПРОСТРАНСТВА

Стаття присвячена розробці логіко-алгебраїчних моделей та методів для формалізації різних семантичних структур природної мови на базі поняття морфологічного простору. З метою удосконалення математичної бази штучного інтелекту ставиться задача формального опису математичних понять на мові алгебри предикатів та предикатних операції, в тому числі й понять, що вводяться самою логічною математикою.

The development of algebrologycal models and methods for formalization of various semantic structures of a natural language is considered. The concept of morphological space was proposed. The task of formal description of mathematical concepts on language of predicates algebra and predicate operations was set in order to refine mathematical base of an artificial intellect.

Постановка проблемы. Невозможно построить полноценный искусственный интеллект пока не будут формализованы основные понятия. Особое значение имеет описание математических понятий. А описание понятий логической математики составляет одну из важнейших задач самой логической математики [1, 2].

Формальное описание объектов дает неизмеримо более полное их описание, иногда даже, исчерпывающее описание. Если мы можем по этому описанию воспроизвести этот процесс так, что искусственная система будет вести себя неотличимо от изучаемой естественной, значит, мы получили исчерпывающее описание данного процесса. Фактически исследователь прибегает к огрублению, схематизации изучаемых процессов и явлений. Такое огрубление приводит к понятию объекта или процесса. Формально каждое описываемое понятие заменяется соответствующим ему предикатом и описывается уже этот предикат [3].

Одно из важнейших практических применений логической математики состоит в формальном описании на ее языке математических понятий, в том числе и понятий, вводимых самой логической математикой.

Анализ литературы. Данная работа является логическим продолжением исследований, опубликованных в работах [4 – 7]. В [6] была предложена идея логической сети – абстрактной структуры, позволяющей обрабатывать дискретную информацию параллельно [8]. В результате исследований по алгебраизации логики была разработана специальная алгебра для формульного представления отношений и действий над ними. Отношения интерпретируются как мысли интеллекта, а действия над ними – как мышление. Схемная реализация формул этой алгебры приводит к характерным инженерным сетям, которые называются логическими сетями.

Целью статьи является разработка аппарата алгебры конечных предикатов для описания структуры обобщенных морфологических пространств. Рассмотреть вопросы формализации различных соответствий, предложить варианты абсолютного и неабсолютного, однозначного и неоднозначного предикатных определений некоторых необходимых понятий. Дать определения проекционного предиката и проектора. На морфологических примерах показать координатные представления векторов в обобщенных пространствах. Предложить методы построения проекционного предиката и сопровождающих квазитолерантностей для заданного пространства. Рассмотреть метод преобразования предикатов к экономной скобочной форме.

1. Понятия как предикаты. Определим формально понятие равенства двух предикатов. Равенство предикатов $P = Q$ представляет собой двуместный предикат второй ступени $=(P,Q)$, аргументами которого являются предикаты первой ступени P и Q. Пусть P и Q – предикаты $P(x_1, x_2, \ldots, x_m)$ и $Q(x_1, x_2, \ldots, x_m)$, заданные на U^m. Сокращенно записываем $(x_1, x_2, \ldots, x_m) = x$, тогда: $=(P,Q) \Leftrightarrow \forall x (P(x) \sim Q(x))$.

Прямым определением предиката Q через предикаты $P_1, P_2, \ldots, P_n$ называется связь вида:

$$\forall x (Q(x) \sim F(P_1(x), P_2(x), \ldots, P_n(x))) = 1, \; (x_1, x_2, \ldots, x_m) = x. \qquad (1)$$

Здесь $P_1, P_2, \ldots, P_n$ – известные предикаты. Они называются параметрами логического уравнения (условия). Q – определяемый предикат; F – некоторая предикатная операция. Уравнение (1) имеет единственное решение, и оно решается в явном виде: предикат Q выражается прямо через предикаты $P_1, P_2, \ldots, P_n$: $Q = F(P_1, P_2, \ldots, P_n)$, или в развернутой форме: $Q(x) \overset{x}{=} F(P_1(x), P_2(x), \ldots, P_n(x))$. Знак $\overset{x}{=}$ означает равенство предикатов (значения предикатов равны при всех наборах значений аргументов x).

Определим формально понятие области определения соответствия. Соответствие – это некоторое отношение, заданное на $M \times N$. M – область отправления, N – область прибытия. Определяем понятие соответствия предикатом $P(x, y)$ на $M \times N$. Область определения соответствия: $A(x) = \exists y \in N \; F(x, y)$. Предикат $A(x)$ задан на M. В полной записи это определение запишется в виде: $\forall x (A(x) \sim M(x) \wedge \exists y \in N \; F(x, y))$; M, N, F – параметры условия (определяющие понятия); A – определяемое понятие. Оно запишется в виде $A(x) \overset{x}{=} M(x) \wedge \exists y \in F(x, y)$.

Определим формально понятие области значений соответствия:

$$\forall y(B(y) \sim N(y) \wedge \exists x \in M\ F(x,y));\ B(y) \overset{x}{=} N(y) \wedge \exists x \in M\ F(x,y);\ B = F(M,N,F).$$

Косвенным определением предиката Q через предикаты $P_1, P_2, \ldots, P_n$ называется связь вида

$$F(P_1(x), P_2(x), \ldots, P_n(x), Q_n(x)) = 1. \qquad (2)$$

Единица, стоящая справа от знака равенства, играет в логических уравнениях роль аналога нуля числовых уравнений. В более общем случае имеем: $P(P_1, P_2, \ldots, P_n,\ Q_1, Q_2, \ldots, Q_r) = 1$, где P – предикат второго порядка.

Прямое определение с дополнительными условиями: E – эквивалентность, F – функция. Прямое определение является частным случаем косвенного. При прямом определении предикат Q всегда выражается однозначно через определяющие его предикаты $P_1, P_2, \ldots, P_n$. При косвенном – не обязательно. Q может не существовать, или же может существовать много вариантов решения уравнения (2) для Q. Непрямое определение – это такое косвенное определение, которое не является прямым. Не исключен случай, когда непрямое определение даст единственный предикат Q. Примером является аксиоматическое определение сложения натуральных чисел: определение непрямое, но оно определяется единственным образом.

Определим формально понятия всюду определенного соответствия. В частном случае, когда $A = M$, имеем: $\forall x(M(x) \approx \wedge \exists y \in N\ F(x,y))$.

Запишем формально определение сюръективного соответствия:

$$\forall y \in N\ \exists x \in M\ F(x,y).$$

Определение предиката Q называется однозначным или коэкстенсивным, если задающее его условие (то есть уравнение алгебры предикатных операций) имеет единственное решение относительно переменной Q. В противном случае оно называется неоднозначным. Все прямые определения однозначны.

Существуют непрямые однозначные определения, например, определение сложения натуральных чисел (рекурсивное). Так что нельзя отождествить прямые определения с однозначными, а непрямые – с неоднозначными.

Определение предиката Q называется абсолютным, если предикат Q не зависит от параметров $P_1, P_2, \ldots, P_n$. В противном случае определение предиката Q называется неабсолютным. Абсолютное определение характеризуется условием $F(Q) = 1$, где F – некоторая предикатная операция.

В частном случае $P(Q) = 1$, где P – некоторый предикат второй ступени. Обычно для записи условий используются замкнутые кванторные выражения, но можно, в принципе, использовать и незамкнутые: $P \supset Q = 1 \Leftrightarrow (P \Rightarrow Q)$; $P \sim Q = 1 \Leftrightarrow (P \Leftrightarrow Q)$. Предикат 1 играет в логических уравнениях такую же роль (правой части уравнения (=1)), как 0 – в уравнениях классической (числовой) математики (=0).

Пример абсолютного неоднозначного определения: $\forall x P(x, x)$, рефлексивность предиката P, заданного на фиксированном универсуме. $\forall x,\ y \in M(P(x, y) \supset P(y, x))$ – симметричность предиката, заданного на фиксированном множестве M.

Пример неабсолютного неоднозначного определения: $\forall x \in M\ P(x, x)$, где $M \subseteq U$ – произвольное подмножество фиксированного универсума U.

2. Построение проекционных предикатов морфологического пространства. Возьмем предикат $S(x_1, x_2, \ldots, x_n, y)$ на $A_1 \times A_2 \times \ldots \times A_n \times B$. Предикат S задает над множеством B пространство S в координатной системе $A = A_1 \times A_2 \times \ldots \times A_n$. Число n называется размерностью пространства S. Множества $A_1 \times A_2 \times \ldots \times A_n$ называются координатными осями пространства S. Предикат S называется характеристикой пространства S. Множество B называется носителем пространства S. Элементы множества B называются векторами или точками пространства S. Элементы $x_1 \in A_1$, $x_2 \in A_2$, …, $x_n \in A_n$, удовлетворяющие условию

$$S(x_1, x_2, \ldots, x_n, y) = 1, \tag{3}$$

называются координатами точки $y \in B$ в системе A. Набор $(x_1, x_2, \ldots, x_n)$, удовлетворяющий условию (3), называется координатным представлением точки *y*. Отображение $S(x_1, x_2, \ldots, x_n) = y$, A в B, соответствующее предикату S, называется отображением координатной системы A пространства S в его носитель B. Предикат $S(x_1, x_2, \ldots, x_n, y)$ связывает каждую точку $y \in B$ с набором $(x_1, x_2, \ldots, x_n)$ ее координат $x_1 \in A_1$, $x_2 \in A_2$, …, $x_n \in A_n$ (одним, многими или ни одним). Отображение $S(x_1, x_2, \ldots, x_n) = y$ каждому набору координат $(x_1, x_2, \ldots, x_n) \in A$ ставит в соответствие точку $y \in B$ (одну, много или ни одной).

i-м проекционным предикатом пространства S называется предикат $G_i(y, x_i)$ на $B \times A_i$ $(i = \overline{1, n})$, значения которого при любых $y \in B$ и $x_i \in A_i$ определяются равенством

$$G_i(y, x_i) = \exists x_1 \in A_1 \exists x_2 \in A_2 .. \exists x_{i-1} \in A_{i-1} \exists x_{i+1} \in A_{i+1} .. \exists x_n \in A_n \quad S(x_1, x_2, \ldots, x_i, \ldots, x_n, y).$$

Предикату $G_i(y, x_i)$ соответствует отображение $g_i(y) = x_i$ из B в A_i, называемое *i*-м проектором пространства S над B. Проекционный предикат $G_i(y, x_i)$ связывает каждую точку $y \in B$ пространства S с ее координатой $x_i \in A_i$ (одной, многими или ни одной). Проектор $g_i(y) = x_i$ каждой точке

$y \in B$ ставит в соответствие ее координату $x_i \in A_i$ (одну, много или ни одной). Иногда, в зависимости от применений, вектор $y \in B$ пространства A называют его предметом, а переменные $(x_1, x_2, \ldots, x_n)$ – признаками предметов; значения переменной x_i $(i = \overline{1, n})$ называют оттенками i-го признака; множество A_i – i-м полем оттенков.

Например, словоформы *большого*, *большому* и т.д. – это предметы. Признаки: слово *большой*, род – (мужской, женский, средний), число (единственное, множественное), падеж (именительный, родительный, дательный, винительный, творительный, предложный), оттенки – мужской и т.п. В классической математике точки плоскости отождествляют с их координатными представлениями в силу наличия биективной связи между ними. Здесь же этого делать нельзя. *Дорогой дядя, дорогой тете, дорогой тетей, на дорогой тете* – разные координатные представления у вектора. *Сук → сучья, суки* – разные векторы соответствуют одному и тому же координатному представлению.

3. Экономные скобочные формы. Скобочными формами называются произвольные формулы дизъюнктивно-конъюнктивной алгебры предикатов. Многие информационные объекты имеют иерархическую структуру, которая отражается при записи предиката в виде скобочной формы. Переход к скобочным формам позволяет до предела уменьшить длину формулы предиката. Для хранения в памяти машины информации о предикате нет ничего более экономного, чем его скобочная форма.

Рассмотрим преобразование к экономной скобочной форме предиката окончаний полных непритяжательных имен прилагательных. Отправляемся от таблицы предиката (т.е. от его СДНФ). Всего имеется 26 трехбуквенных окончаний, т.о. СДНФ предиката имеет 26×3=78 предикатов узнавания предмета. Такова сложность исходной формулы. Сложность формулы оцениваем числом вхождений в нее предикатов узнавания предмета. Все окончания парные, в каждой паре последние две буквы окончания одинаковы. Учитывая это обстоятельство, переходим к скобочной форме:

$$P(x_1, x_2, x_3) = (x_1^{\text{ы}} \vee x_1^{\text{и}})((x_2^{\text{й}} x_3^{-} \vee x_2^{\text{м}} x_3^{-} \vee x_2^{\text{е}} x_3^{-} \vee x_2^{\text{х}} x_3^{-} \vee x_2^{\text{м}} x_3^{\text{и}}) \vee$$

$$\vee (x_1^{\text{о}} \vee x_1^{\text{е}})(x_2^{\text{г}} x_3^{\text{о}} \vee x_2^{\text{м}} x_3^{\text{у}} \vee (x_2^{\text{м}} x_3^{-} \vee x_2^{\text{й}} x_3^{-} \vee x_2^{\text{ю}} x_3^{-} \vee x_2^{\text{е}} x_3^{-}) \vee$$

$$\vee (x_1^{\text{а}} \vee x_1^{\text{я}}) x_2^{\text{я}} x_3^{-} \vee (x_1^{\text{у}} \vee x_1^{\text{ю}}) x_2^{\text{ю}} x_3^{-}.$$

Число предикатов узнавания предмета сократилось с 78 до 32, т.е. получили более, чем двукратную экономию. Это – огромная экономия. Но это еще не предел, т.к. в таблице имеются глубокие закономерности. Можно вынести за скобки еще x_3^{-}, где только возможно:

$$P(x_1, x_2, x_3) = (x_1^{\text{ы}} \vee x_1^{\text{и}})((x_2^{\text{й}} \vee x_2^{\text{м}} \vee x_2^{\text{е}} \vee x_2^{\text{х}})x_3^{-} \vee x_2^{\text{м}} x_3^{\text{и}}) \vee$$

$$\vee (x_1^{\text{о}} \vee x_1^{\text{е}})(x_2^{\text{г}} x_3^{\text{о}} \vee x_2^{\text{м}} x_3^{\text{у}} \vee (x_2^{\text{м}} \vee x_2^{\text{й}} \vee x_2^{\text{ю}} \vee x_2^{\text{е}})x_3^{-}) \vee (x_1^{\text{а}} \vee x_1^{\text{я}})x_2^{\text{я}} \vee (x_1^{\text{у}} \vee x_1^{\text{ю}})x_2^{\text{ю}})x_3^{-}.$$

Для каждой из них в формуле предиката должно фигурировать хотя бы одно вхождение предиката узнавания. Дальнейшее сокращение формулы возможно за счет выделения в ней общих частей. Имеется только одна общая часть $A = x_2^{\text{м}} \vee x_2^{\text{й}} \vee x_3^{\text{е}}$. Заменяем ее именем A.

$$P(x_1, x_2, x_3) = (x_1^{\text{ы}} \vee x_1^{\text{и}})((A \vee x_2^{\text{х}})x_3^{-} \vee x_2^{\text{м}} x_3^{\text{и}}) \vee$$

$$\vee (x_1^{\text{о}} \vee x_1^{\text{е}})(x_2^{\text{г}} x_3^{\text{о}} \vee x_2^{\text{м}} x_3^{\text{и}} \vee (A \vee x_2^{\text{ю}})x_3^{-}) \vee (x_1^{\text{а}} \vee x_1^{\text{я}})x_2^{\text{я}} \vee (x_1^{\text{у}} \vee x_1^{\text{ю}})x_2^{\text{ю}})x_3^{-}.$$

Окончательный результат в сокращенном виде:

$$(\text{ы} \vee \text{и})((\text{А} \vee \text{х}) \vee \text{ми}(\text{о} \vee \text{е})\ (\text{го} \vee \text{ми})(\text{А} \vee \text{ю})) \vee (\text{а} \vee \text{я})\text{я} \vee (\text{у} \vee \text{ю})\text{ю};$$

$$\text{А} = \text{м} \vee \text{й} \vee \text{е}.$$

Выводы. В результате работы формализован ряд основных понятий алгебры конечных предикатов, что позволило ввести структуру обобщенных морфологических пространств. Даны и проиллюстрированы на примерах важные определения морфологического пространства, в частности, проекционного предиката и проектора. Рассмотрен метод преобразования предикатов к экономной скобочной форме. Перспективы дальнейших исследований в данном направлении определяются эффективностью полученных методов при параллельной обработке логических связей и возможностью моделирования любых лингвистических объектов в виде логических сетей.

Список литературы: **1.** *Дударь З.В., Пославский С.А. и др.* Предикаты эквивалентности в задачах компараторной идентификации // Проблемы бионики. – 1999. – Вып. 51. – С. 19–26. **2.** *Дударь З.В., Мельникова Р.В., Шабанов–Кушнаренко Ю.П.* Отношения как предмет формульного описания // Проблемы бионики. – 1997. – Вып. 48. – С. 94–103. **3.** *Шабанов-Кушнаренко Ю.П.* Теория интеллекта. Математические средства. – Харьков: “Вища школа”, 1984. – 142 с. **4.** *Бондаренко М.Ф.* Математические модели морфологических отношений и их применение для автоматизации обработки речевых сообщений. – Дис. ... докт. техн. наук: 05.13.01. – Харьков, 1984.– 349 с. **5.** *Шабанов–Кушнаренко Ю.П.* Теория интеллекта – Ч.2: Технические средства. – Харьков: Вища школа, 1986. – 144 с. **6.** *Бондаренко М.Ф., Дударь З.В., Ефимова И.А., Лещинский В.А., Шабанов-Кушнаренко С.Ю.* О мозгоподобных ЭВМ. // Радиоэлектроника и информатика. – 2004. – № 3 (04). – С. 24–32. **7.** *Дударь З.В., Калиниченко О.В., Козяев Л.Л., Мельникова Р.В.* О формальной структуре логических сетей // Радиоэлектроника и информатика. – 2005. – № 2. – С. 45–51. **8.** *Козяев Л.Л., Шабанов-Кушнаренко С.Ю.* Бинаризация морфологического отношения склонения полных имен прилагательных // Радиоэлектроника и информатика. – 2005. – № 2. – С. 36–44.

Поступила в редакцию 17.03.2005