

УДК 62.50

Н. Г. ЛЫСЕНКО, Г. Ф. КРИВУЛЯ, канд. техн. наук

**КОДИРОВАНИЕ СИМВОЛОВ РАСПОЗНАЮЩЕЙ ГРАММАТИКИ
ДВУХОСНОВНОЙ АЛГЕБРЫ ПРЕДИКАТА**

Все методы кодирования связаны с решением хотя бы одной из трех задач: передачей, хранением и преобразованием информации. В данной статье рассматриваются только известные методы кодирования хранимой и преобразуемой информации [1] применительно к символам распознающей грамматики, порождаемой двухосновной алгеброй предиката [2],

поскольку в рамках этой грамматики вопросы теории кодов, исправляющих ошибки [3], не выходят за пределы задачи распознавания образов [2]. В существующих методах оценки характеристик кода его емкость задается эвристически, что приводит к субъективности и, следовательно, к недолговечности классификатора. Отличительная особенность кодирования символов грамматики двухосновной алгебры предиката состоит в возможности более объективно оценивать емкость кода.

Методы кодирования символов. Наиболее часто терминальные символы кодируют дескрипторно [1] из-за простоты кодирования новых свойств, а нетерминальные — последовательно блоками [1], основываясь на использовании иерархической классификации свойств последних. Поскольку в рассматриваемой грамматике синтаксический анализ разверткой ведется методом группового учета аргументов (МГУА) по вертикали и МГУА по горизонтали [2], наиболее целесообразно кодировать символы этими аргументами. Последовательное кодирование символов объединенного алфавита (как, например, в [2]) дает наглядность представления дерева грамматики. Программная реализация схемы грамматики не требует такой наглядности, что позволяет отказаться от последовательного кодирования в пользу параллельного кодирования аргументами по вертикали, обеспечивая простоту МГУА. Поскольку все аргументы нетерминального символа являются параметрами его грубой модели, возможна многоуровневая классификация символов (нетерминальный символ может быть представлен различными комбинациями), позволяющая ускорить поиск.

Каждый уровень нетерминальных символов (предложение, слово, слог) параллельно кодируется только одним аргументом анализа по вертикали на первых разрядах с порядковой регистрацией группировок нетерминальных символов на остальных разрядах, отличающихся символами следующего, более низкого уровня иерархии. Физический смысл порядковой регистрации группировок символов состоит также и в указании адреса области памяти классификатора с хранимыми параметрами анализа по горизонтали. Необходимость хранения таких параметров состоит в том, чтобы при расширении распознающей грамматики при включении в нее все новых порождающих грамматик не кодировать заново встречающиеся ранее символы. Поскольку эти параметры для терминальных символов независимы, их можно рассматривать как отдельные признаки параллельного кодирования блоками, как, например, в терминах описания массивов на языке программирования ПЛ1 [2].

Применение идеи грубой модели к нетерминальным символам позволяет отказаться от их последовательного кодирования и использовать параллельное кодирование как для терминальных символов, обеспечивая тем самым идентичность и простоту обработки по отдельным признакам. Физический

смысл параллельного кодирования нетерминальных символов заключается в указании их габаритов, а терминальных — в указании их размеров. Полный отказ от параллельно-дескрипторного в пользу параллельного кодирования блоками значительно увеличил бы объем памяти, отводимой под код вершин деревьев вывода. Компромиссом является хранение в одном экземпляре каждого символа в параллельном коде блоками и в необходимом количестве — в параллельно-дескрипторном коде в вершинах деревьев вывода. Связь между «внешним» и «внутренним» кодом символов представляет собой систему ассоциаций [2].

В отличие от синтаксического анализ образов предполагает заведомо их знание. В таком случае кодирование символов особенно претерпевает существенные изменения при снятии ограничений с объема памяти для хранения табличных моделей символов (например, когда символов немного). При этом вместо параллельно-дескрипторного кода в вершинах деревьев вывода можно использовать непосредственно параллельный код и отказаться от системы ассоциаций.

Вычисление характеристик кода. Емкость кода символов рассматриваемой грамматики определяется размерами координатной сетки, накладываемой на объект. В свою очередь, размер N координатной сетки определяется допустимой ошибкой ϵ в долях аппроксимации минимального символа по размерам, допустимой ошибкой φ в радианах аппроксимации минимального символа, по углу наклона и отношению o габаритов максимального нетерминального символа к размерам минимального терминального символа (предполагается для простоты идентичность требований по обеим координатам):

$$N = \frac{o}{\epsilon \varphi}.$$

Максимальное число M типов терминальных символов при этом

$$M = \frac{N(N+1)}{2},$$

а терминальных символов в слове равно N , слогов в слове — $0,5N^2$, слов в предложении — $0,5N^2$. Таким образом, верхняя оценка числа R разрядов под каждый признак параллельного кода терминального символа с основанием a равна

$$R = \log_a N,$$

как и число разрядов под признак дескрипторной части параллельно-дескрипторного кода слога и параллельной части кодов всех нетерминальных символов.

Верхняя оценка числа разрядов R_- под признак дескрипторной части параллельно-дескрипторного кода слова и предложения

$$R_- = 2 \log_a (0,5N).$$

Так, размер координатной сетки при $\epsilon = 0,1$, $\varphi = 0,1$ и $\sigma = 10$ равен 1000. Максимальное число терминальных символов, размещаемых на этой координатной сетке, равно 5×10^5 . Каждый из пяти [2] признаков параллельного кода терминального символа представляется тремя цифрами, а дескрипторная часть параллельно-дескрипторного кода состоит из шести цифр. Общее число символов алфавита равно примерно 10^6 . Максимально допустимое количество символов в алфавите растет в квадратичной зависимости от размера координатной сетки, тогда как число всевозможных образов растет значительно быстрее (как функция 2^{N^2}), так что предлагаемое грамматическое описание терминальными и нетерминальными символами и их кодирование позволяют существенно минимизировать табличное описание объекта.

Список литературы: 1. Касами Т., Токура Н., Ивадари Е., Инагаки Я. Теория кодирования. М.: Мир, 1978.— 576 с. 2. Лысенко Н. Г., Кривуля Г. Ф. Синтаксис и семантика порождающей и распознающей грамматик двухосновной алгебры предиката // АСУ и приборы автоматики. — 1985.— Вып. 75.— С. 17—21. 3. Венчковский Л. Б. Методы кодирования технико-экономической информации.— М.: Сов. радио, 1978.— 120 с.

Поступила в редколлегию 07.07.83