УДК 621.3 + 681.3

Д.В. СПЕРАНСКИЙ

Саратовский государственный университет, Россия

ЭВОЛЮЦИОННЫЕ МЕТОДЫ В ЗАДАЧАХ СИНТЕЗА ТЕСТОВ ДЛЯ ДИСКРЕТНЫХ УСТРОЙСТВ[*/]

Предложен генетический алгоритм для решения задачи синтеза тестов для линейных автоматов, являющихся математическими моделями широко распространенных на практике дискретных устройств. Задача решается в предположении, что реакции автоматов содержат неопределенности, выраженные в виде интервальных значений.

дискретные устройства, линейные автоматы, тесты, генетические алгоритмы

Введение

На практике мы часто имеем дело с дискретными устройствами, которые используются для кодирования и декодирования информации, сжатия информации, для шифрации и дешифрации сообщений и т.п. Все эти устройства имеет общую особенность: они адекватно описываются моделями линейных автоматов (ЛА), которые заданы над конечным полем $GF(p) = \{0, 1, \ldots, p-1\}$, где p – простое число. Классическая теория ЛА [1] была развита в предположении, что все параметры ЛА и наблюдаемые реакции являются точными значениями. В то же время очевидно, что такое предположение не всегда справедливо. В действительности мы часто имеем дело с информацией, содержащей некоторую неопределенность. В частности, такая неопределенность может возникать из-за ошибок измерений. Если неопределенность имеет место, то мы достоверно можем указать только интервал, который содержит истинное значение измеряемого параметра. Таким образом, мы приходим к необходимости использования интервальной арифметики и соответствующих операций над конечным полем $GF(p)$ и разработке методов нахождения решений для смешанных уравнений с точными коэффициентами и интервальной правой частью. Заметим, что для непрерывных систем над полем вещественных чисел соответствующий математический аппарат (интервальный анализ) достаточно хорошо развит [2]. Основы интервальной арифметики над конечными полями представлены в [3].

Поскольку концепция интервала для конечного поля отличается от концепции интервала в поле вещественных чисел, мы рассмотрим это чуть подробнее.

Каждый элемент поля $GF(p)$ представляет собой класс вычетов по модулю p. Выберем одного представителя из каждого такого класса, который является минимальным неотрицательным числом. Далее мы введем такой же порядок на множестве всех упомянутых представителей, как порядок на множестве целых чисел. Используемое выше обозначение $GF(p)$ будет сохранено и для упорядоченного множества $[0,1,\ldots,p-1]$.

Подмножество $\boldsymbol{a} \subseteq GF(p)$ такое, что $\boldsymbol{a} = [\underline{a}, \overline{a}] = \{\, a \mid \underline{a} \le \alpha \le \overline{a};\, \overline{a}, \underline{a} \in GF(p)\}$ мы назовем правильным интервалом, где $\overline{a}$ и $\underline{a}$ есть верхняя и нижняя границы интервала. Интервал $\boldsymbol{b} = [\underline{b}, \overline{b}] = GF(p) \setminus [\overline{b}+1, \underline{b}-1]$ будем называть неправильным интервалом, где знак \ есть знак теоретико-множественного вычитания. Интервал $[\underline{a}, \overline{a}]$, где $\underline{a} = \overline{a}$ будем называть вырожденным интервалом и интерпретировать как элемент поля $GF(p)$.

Далее все упомянутые выше типы интервалов будем называть просто интервалами и определим арифметические операции над ними.

Пусть $* \in \{+,-,\cdot,/\}$ есть бинарная операция над элементами поля $GF(p)$. Если $\boldsymbol{a}$, $\boldsymbol{b}$ есть интервалы, тогда выражение $\boldsymbol{a} * \boldsymbol{b} = \{\alpha * \beta \mid \alpha \in \boldsymbol{a}, \beta \in \boldsymbol{b}\}$ определяет арифметическую операцию $*$ над интервалами. Для операции деления предполагается, что $0 \notin \boldsymbol{b}$.

Заметим, что в отличие от поля вещественных чисел результатом операций над интервалами в поле

[*)] Работа поддержана грантами РФФИ 05-08-18082, 05-08-49999

$GF(p)$ может быть объединение нескольких интервалов.

Объект исследования и постановка задачи. Объектом исследования является ЛА над полем ***GF*(*p*)**, который задан уравнениями переходов и выходов:

$$\bar{s}(t+1) = A\bar{s}(t) + B\bar{u}(t), \quad (1)$$

$$\bar{y}(t) = C\bar{s}(t) + D\bar{u}(t), \quad (2)$$

где A, B, C, D – характеристические матрицы размерности $n \times n, n \times l, m \times n, m \times l$ соответственно. Векторы $\bar{u}(t) = [u_1(t), \ldots, u_l(t)]'$, $\bar{y}(t) = [y_1(t), \ldots, y_m(t)]'$, $\bar{s}(t) = [s_1(t), \ldots, s_n(t)]'$ представляют входы, выходы и состояния ЛА соответственно.

Ниже исследуется следующая задача. Пусть задан ЛА A и множество модификаций этого автомата $M = \{A_1, \ldots, A_k\}$. Требуется построить тест, который позволяет для проверенного ЛА ответить на вопрос: является ли он исправным либо принадлежит множеству автоматов M .

При этом предполагается, что наблюдаемая реакция проверяемого ЛА на тест есть последовательность выходных векторов с интервальными координатами.

Входную последовательность мы назовем тестом, если независимо от начального состояния исправного ЛА и любого автомата из множества M их выходные реакции различны.

Пусть $\bar{y}_u(q)$ есть точное значение реакции исправного ЛА на некоторый входной символ, которое вычислено по формуле (2). Пусть $\pm\Delta$ есть точность, обеспечиваемая используемым измерительным прибором, которая всегда известна. Тогда интервал $[\bar{y}_u(q) - \Delta, \bar{y}_u(q) + \Delta]$ содержит истинное значение выхода исправного ЛА в момент времени q . Пусть $\bar{y}_{np}(q)$ есть интервальная реакция проверяемого ЛА в момент времени q . Будем считать эти две интервальные реакции различными, если интервалы представляющие их соответствующие координаты, не пересекаются (т.е. их пересечение пусто).

Метод синтеза теста с применением генетического алгоритма

Сформулированная задача синтеза теста может быть решена различными известными методами, включая точные аналитические, случайные и псевдослучайные. Однако все эти методы для ЛА большой размерности являются весьма трудоемкими и не дают гарантии построения теста. По этой причине ниже предлагается метод синтеза тестов с помощью генетического алгоритма (ГА) [4], хорошо зарекомендовавшего себя для решения сложных оптимизационных проблем большой размерности. Все термины и понятия, используемые ниже, касающиеся ГА, трактуются так, как это приведено в монографии [4].

ГА является адаптивным поисковым методом, базирующемся на отборе «лучших» особей в популяции, то есть наиболее приспособленных к окружающей среде, как это происходит в дарвиновской теории эволюции. ГА эффективно использует информацию, которая накапливается в процессе эволюции.

Для использования ГА необходимо выбрать множество параметров оптимизационной задачи и закодировать их в виде битовых строк конечной длины.

Для завершения работы ГА используются различные критерии: число сгенерированных популяций превысило заданный предел; получены особи определенного качества; найден локальный оптимум.

Все ГА работают на основе начальной информации, которая представляет собой альтернативные решения $P = \{p_1, \ldots, p_N\}$, которые образуют начальную популяцию размерности N . Каждый элемент популяции p_i обычно представляет собой хромосому (особь). Хромосомы состоят из генов.

Процесс работы ГА базируется на множестве правил, таких как выбор в популяции родительских пар, которые далее скрещиваются для получения потомства. Новая популяция образуется из полученного потомства и родительских пар. Результат реализации одной итерации называется генерацией.

Эволюция популяции есть чередование генераций, в которой хромосомы изменяются так, что каждая новая генерация лучше приспособлена к окружающей среде. Каждая особь в популяции имеет определенную степень качества, которая характеризуется значением фитнес-функции. Для решения задачи с применением ГА необходимо выполнить следующие этапы: 1) выбрать форму представления решения (структуру хромосом); 2) создать начальную популяцию; 3) определить операторы случайных изменений; 4) определить законы выживания особей (решений).

Возвращаясь к исследуемой задаче отметим, что при использовании ГА для синтеза теста начальной информацией является исправных ЛА и множество M его неисправных модификаций.

Искомый тест представляет собой входную последовательность T, которая интерпретируется как хромосома. Геном в хромосоме является входной вектор $\overline{u}(t)=[u_1(t),...,u_l(t)]'$, где $u_i \in \{0,1\}$, т.е. каждый ген в хромосоме имеет одну и ту же длину.

Важным моментом для реализации ГА является выбор подходящей фитнес-функции, которая используется для оценки приспособленности хромосомы к окружающей среде. Для вычисления значения фитнес-функции для конкретной хромосомы необходимо преобразовать (декодировать) эту хромосому в соответствующее решение. В качестве фитнес-функции мы будем использовать функцию

$$f(z)=-\sum_{i=1}^{k}\frac{\left|M^{(i)}\right|\circ\lambda_i}{|M|}, \qquad (3)$$

где k – длина теста; z – хромосома; $\left|M^{(i)}\right|$ – число ЛА в множестве M, которые на первых $(i-1)$ входных символах теста дают реакции, совпадающие с реакцией исправного ЛА, а на i-м символе реакции отличаются. Величина $\lambda_i=|M|-i+1$ есть коэффициент, обеспечивающий ухудшение «качества» теста с ростом его длины.

Из (3) ясно, что чем меньше длина теста, тем больше значение фитнес-функции. Очевидно, что это соответствует здравому смыслу применительно к тестам: чем короче тест, тем он предпочтительнее.

Для апробации стандартного простого ГА применительно к рассматриваемой задаче синтеза теста была использована система Mat Lab 7.0, которая осуществляет только минимизацию фитнес-функции. По этой причине в формуле (3) перед суммой стоит знак «минус». Заметим, что начальная популяция в используемом варианте ГА формируется случайным образом с помощью соответствующего датчика случайных чисел.

Для исследования предложенного ГА были проведены численные эксперименты, позволившие определить влияние различных параметров на эффективность ГА. В результате проведенных экспериментов было установлено, что элитный принцип отбора особей в очередную популяцию иногда отрицательно сказывается на длине теста и что мутация почти не сказывается на длине теста.

В предложенном варианте ГА используемый оператор кроссовера отличается от стандартного, что связано с необходимостью попадания точки разрыва на конец входного символа теста, ибо его разрыв недопустим. Тестирование с использованием такого одноточечного кроссовера дало хорошие результаты.

Заключение

Результаты численных экспериментов дают основание утверждать, что предложенный вариант ГА для решения задачи синтеза тестов в приведенной выше формулировке путем подбора параметров ГА, обеспечивающих адаптацию алгоритма к конкретному ЛА и множеству M его неисправных модификаций, может обеспечивать достаточно высокую эффективность синтеза при приемлемой длине теста.

Литература

1. Гилл А. Линейные последовательностные машины (анализ, синтез и приложения). – М.: Наука, 1974.

2. Алефельд Г., Херцбергер Дж. Введение в интервальные вычисления. – М.: Мир, 1985.

3. Куприянова Л.В., Сперанский Д.В., Самойлов В.Г. Интервальная арифметика над полем *GF*(*p*) // Вычислительные технологии. – 2002. – Т. 7, № 6. – С. 54–64.

4. Емельянов В.В., Курейчик В.М., Курейчик В.В. Теория и практика эволюционного моделирования. – М.: Физматиз, 2003.

Поступила в редакцию 24.01.2006

Рецензент: д-р техн. наук, проф. В.С. Харченко, Национальный аэрокосмический университет им. Н.Е. Жуковского «ХАИ», Харьков.