АЛГОРИТМ ФОРМИРОВАНИЯ УПРАВЛЯЕМЫХ ВЕРОЯТНОСТНЫХ ТЕСТОВ С ЗАДАННЫМ РАССТОЯНИЕМ ХЭММИНГА

Леванцевич В. А., Деменковец Д. В.
Кафедра программного обеспечения информационных технологий,
Белорусский государственный университет информатики и радиоэлекторoники
Минск, Республика Беларусь
E-mail: lvn@blsuir.by, Demenkovets@bsuir.by

Рассматриваются управляемые вероятностные тесты, методы их формирования, а также их применение для тестирования вычислительных систем. Анализируются численные метрики для построения управляемых вероятностных тестов, Показываются основные проблемы, возникающие при формировании таких тестов. Рассматривается алгоритм формирования управляемых вероятностных тестов, основанный на определении расстояния Хэмминга для тестовых наборов, состоящих из символов различных алфавитов.

Введение

Вероятностное тестирование (*Random Testing*) и его многочисленные модификации, основанные на принципе *черного ящика* (*Black Box*), являются эффективным средством для тестирования вычислительных систем.

Для повышения эффективности вероятностных тестов используют их модификации, которые получили общее название управляемые вероятностные тесты (*Controlled Random Tests*). Общим для управляемых вероятностных тестов является то, что каждый последующий тестовый набор формируется с учетом ранее сгенерированных наборов на основе вычисления некоторых характеристик, по которым он включается или не включается в вероятностный тест. Основной проблемой данных тестов является высокая вычислительная сложность определения очередного кандидата в тестовый набор. [1].

I. Анализ предметной области

В качестве меры отличия тестовых наборов $T_i = t_{i,0}\ \ t_{i,1}\ \ t_{i,2}\ \ \ldots\ \ t_{i,n-1}$ и $T_j = t_{j,0}\ \ t_{j,1}\ \ t_{j,2}\ \ \ldots\ \ t_{j,n-1}$, где $t_{i,l},\ t_{j,l} \in 0{,}1$, $l \in \{0,\ \ 1,\ \ 2,\ \ \ldots,\ \ n-1\}$, чаще всего используются кодовое расстояние Хемминга. Однако, расстояние Хэмминга как мера различия не всегда эффективна (см таблицу 1). Для более полной оценки различия двоичных наборов в [2-3] была определена новая мера различия $MD(T_i,T_j)$ между тестовыми наборами T_i и T_j для случая, когда $n = 2^m$, где m – целое. Данная мера различия состоит из множества численных характеристик $HD_0,HD_1,\ldots,HD_v,\ldots,HD_m$, представляющих собой расстояния Хэмминга $HD_v[T_i(2^v),T_j(2^v)]$ для указанных наборов $T_i(2^v)$ и $T_j(2^v)$, состоящих из символов различных алфавитов, заданных их 2^v последовательными битами, где $v \in \{0,\ \ 1,\ \ 2,\ \ \ldots,\ \ m\}$.

В табл. 1 приведены примеры вычисления $MD(T_i,T_j)$ для различных пар тестовых наборов T_i и T_j разрядностью $n = 5$.

Таблица 1 – Примеры вычисления расстояния Хэмминга $HD(T_i,T_j)$

T_i	$T_i(1) = 0\,1\,0\,1\,0$	$T_i(2) = 01\,01\,00$
T_j	$T_j(1) = 1\,0\,0\,0\,0$	$T_j(2) = 10\,00\,00$
$HD(T_i,T_j)$	3	2
T_i	$T_i(3) = 010\,100$	$T_i(4) = 0101\,0000$
T_j	$T_j(3) = 100\,000$	$T_j(4) = 1000\,0000$
$HD(T_i,T_j)$	2	1
T_i	$T_i(5) = 01010$	
T_j	$T_j(5) = 10000$	
$HD(T_i,T_j)$	1	
T_i	$T_i(1) = 0\,1\,0\,1\,0$	$T_i(2) = 01\,01\,00$
T_j	$T_j(1) = 1\,0\,0\,1\,1$	$T_j(2) = 10\,01\,10$
$HD(T_i,T_j)$	3	2
T_i	$T_i(3) = 010\,100$	$T_i(4) = 0101\,0000$
T_j	$T_j(3) = 100\,110$	$T_j(4) = 1001\,0000$
$HD(T_i,T_j)$	2	1
T_i	$T_i(5) = 01010$	
T_j	$T_j(5) = 10011$	
$HD(T_i,T_j)$	1	

Как видно из таблицы мера различия $MD(T_i,T_j)$ принимает следующие значения: $MD(01010{,}10000) = HD_1 = 3$, $HD_2 = 2$, $HD_3 = 2$, $HD_4 = 1$, $HD_5 = 1$; $MD(01010{,}10011) = 3{,}2{,}2{,}2{,}1$.

Использование меры различия $MD(T_i,T_j)$ позволяет сформировать управляемый вероятностный тест, состоящий из $q = 2^r$ двоичных наборов разрядностью n, где $r \in \{0,\ \ 1,\ \ 2,\ \ \ldots,\ \ n\}$ есть количество бит используемых для двоичного представления символа определенного алфавита. Тестовые наборы такого теста будут иметь кодовое расстояние, равное $h = minHD(T_i,T_j) = n/r$.

Ограниченное количество $q = 2^r$ тестовых наборов определяется ограниченным количеством 2^r символов алфавита, в котором представлены тестовые наборы $T_i(r) = t_{i,0}(r)\ \ t_{i,1}(r)\ \ t_{i,2}(r)\ \ \ldots\ \ t_{i,n/r-1}(r)$ и $T_j(r) = t_{j,0}(r)\ \ t_{j,1}(r)\ \ t_{j,2}(r)\ \ \ldots\ \ t_{j,n/r-1}(r)$ Только в этом случае символы одних и тех же цифр (разрядов) во всех q тестовых наборах могут принимать разные значения без повторений.

II. Реализация

Для построения управляемого вероятностного теста часто используются исходные шаблоны. Под тестовыми шаблонами подразумевается управляемый вероятностный тест $CRT(q,h,n)$ с фиксированным количеством тестовых наборов q, заданным минимальным значением расстояния Хэмминга $h = minHD(T_i,T_j)$, построенными для минимальной разрядности n тестовых наборов.

Используя подобные шаблоны с заданными характеристиками q и h, строится управляемый вероятностный тест для требуемой разрядности n тестовых наборов, в котором сохраняется относительное h/n значение расстояния Хэмминга.

Процедура построения управляемых вероятностных тестов с фиксированным расстоянием Хэмминга, исходными данными для которой являются разрядность n двоичных тестовых наборов и требуемое значение $minHD(T_i,T_j)$, включает следующие шаги.

1. Определяется максимальное значение $r \in \{0, \ 1, \ 2, \ 3, \ \ldots, \ n\}$, для которого выполняется неравенство $minHD(T_i,T_j) \leq \lfloor n/r \rfloor$. Это значение, согласно утверждению 2, определяет число q тестовых наборов в тесте, равное 2^r, и расстояние Хэмминга $minHD(T_i,T_j) \geq \lfloor n/r \rfloor$ для всех $i \neq j \in \{0, \ 1, \ 3, \ \ldots, \ 2^r - 1\}$.

2. Значения первых r бит 2^r тестовых наборов $T_0,T_1,T_2,\ldots,T_{q-1}$ устанавливаются равными двоичному коду одного из символов алфавита, содержащего 2^r символов. Двоичные коды символов формируются случайным образом без повторений, начиная с первого набора T_0 до последнего T_{q-1}. Для этих целей можно использовать шаблоны из семейства шаблонов $CRT(2^r,1,r)$. Таким образом, каждый тестовый набор будет содержать в первых r разрядах уникальную двоичную комбинацию, соответствующую одному из 2^r символов.

3. Шаг 2 повторяется $\lfloor n/r \rfloor - 1$ раз для всех последующих r разрядных блоков двоичных тестовых наборов. То есть на второй итерации задаются уникальные r –битовые коды следующих r разрядов, а именно $r,r+1,\ldots,2r-1$ бит тестовых наборов.

4. Значения оставшихся $n - \lfloor n/r \rfloor \times r$ разрядов, если таковые имеются, всех тестовых наборов формируются случайным образом. Приведем пример применения рассмотренной процедуры для синтеза управляемого вероятностного теста для $n = 16$ и $minHD(T_i,T_j) = 5$.

На основании неравенства $minHD(T_i,T_j) = 5 \leq \lfloor n/r \rfloor = \lfloor 16/r \rfloor$ получаем значение $r = 3$, так как оно является максимальным значением r из множества его значений $\{1, \ 2, \ 3, \ \ldots, \ 16\}$, при котором выполняется данное неравенство. Соответственно, формируемый тест будет состоять из $2^r = 2^3 = 8$ наборов.

1. Значения первых 3-х разрядов $t_{i,0}$, $t_{i,1}$ и $t_{i,2}$ тестовых наборов T_0, T_1, T_2, $\ldots$, T_7 устанавливаются равными одному из двоичных кодов $000,001,010,\ldots,111$ символов $0,1,2,\ldots,7$ восьмеричного алфавита. Для этого можно использовать один из шаблонов $CRT(2^3,1,3)$, например, 010, 101, 011, 000, 111, 001, 110 и 100. Таким образом, каждый тестовый набор будет содержать в первых 3-х разрядах уникальную двоичную комбинацию (см. табл. 2).

2. Шаг 2 повторяется $\lfloor 16/3 \rfloor - 1 = 4$ раза. Для каждого блока, состоящего из трех бит, значения символов восьмеричного алфавита назначаются случайным образом без повторений.

3. Значения оставшегося $16 - \lfloor 16/3 \rfloor \times 3 = 1$ бита $t_{i,15}$ всех тестовых наборов формируются случайным образом.

В результате применения рассмотренной процедуры получен тест, представленный в табл. 2, для которого выполняется следующее условие: $HD(T_i,T_j) \geq minHD(T_i,T_j) = 5$. Как видно из табл. 2, все значения $HD(T_i,T_j)$ больше или равны 5.

Таблица 2 – Управляемый вероятностный тест с $minHD(T_i,T_j) = 5$ и $n = 16$

T	$t_{i,0}$	$t_{i,1}$	$t_{i,2}$	$t_{i,3}$	$t_{i,4}$	$t_{i,5}$	$t_{i,6}$	$t_{i,7}$
T_0	0	1	0	0	0	0	0	1
T_1	1	0	1	1	1	0	0	1
T_2	0	1	1	1	0	0	1	0
T_3	0	0	0	1	1	1	0	0
T_4	1	1	1	0	0	1	1	1
T_5	0	0	1	0	1	0	0	0
T_6	1	1	0	1	0	1	1	0
T_7	1	0	0	0	1	1	1	1
T	$t_{i,8}$	$t_{i,9}$	$t_{i,10}$	$t_{i,11}$	$t_{i,12}$	$t_{i,13}$	$t_{i,14}$	$t_{i,15}$
T_0	1	1	1	1	0	1	0	0
T_1	0	0	1	1	1	1	1	1
T_2	1	1	0	0	1	0	0	1
T_3	0	0	1	0	0	0	1	1
T_4	0	1	0	1	1	1	0	0
T_5	1	0	0	1	1	0	1	1
T_6	0	0	0	0	0	1	1	0
T_7	1	1	1	0	0	0	0	0

III. Выводы

Предложенный алгоритм позволяет формировать управляемые вероятностные тесты без необходимости перечисления кандидатов в тестовые наборы, что сводит задачу синтеза управляемого вероятностного теста к формальной процедуре, предложенной в настоящей работе.

1. Ярмолик, В. Н. Контроль и диагностика вычислительных систем / В. Н. Ярмолик. – Минск : Бестпринт, – 2019. – 387 с.
2. Ярмолик, В. Н. Мера различия для управляемых вероятностных тестов / В. Н. Ярмолик, В. В. Петровская, Н. А. Шевченко // Доклады БГУИР. – 2024. – № 4(22). – С. 76–83.
3. Ярмолик, В. Н. Модификации способов определения расстояния Хэмминга для их применения в качестве мер различия при генерировании управляемых вероятностных тестов / В. Н. Ярмолик, В. В. Петровская, Н. А. Шевченко // Информатика. – 2024. – Т. 21, № 2. – С. 54–72.