

Лабораторная работа №1.

Вероятностное моделирование метрических характеристик программ

Цель работы

Целью работы является разработка программы, имитирующей процесс написания программы программистом, и расчет статистических оценок метрических характеристик полученных программ. Для имитации используется вероятностная модель выборки с возвратом из генеральной совокупности, состоящей из η символов, пока последняя не будет исчерпана, а также проведение серий испытаний при разных значениях η .

Вывод уравнения длины программы

Условимся считать, что любая программа, написанная программистом на каком-либо языке программирования (будь то ассемблер или язык высокого уровня) состоит из последовательности имен операторов и операндов. Также программа обычно удовлетворяет следующим условиям:

1. Маловероятно появление имени оператора или операнда много раз подряд. Как правило, развитые языки программирования позволяют построить такую конструкцию, где подобные фрагменты программы были бы минимальной длины.
2. Циклическая организация программы исключает многократное появление какой-либо группы операторов и операндов.
3. Блоки программы, требующие многократного повторения при ее исполнении, обычно оформляются как подпрограммы: процедуры или функции, так что в тексте программы присутствуют лишь их имена.
4. Имя каждого оператора и операнда должно появляться в тексте программы хотя бы один раз.

Таким образом, текст программы с циклами можно рассматривать как сжатую последовательность команд развертки этой программы на машинном языке. Согласно определению алгоритмической сложности объекта, если текст не может быть записан более коротким способом, то последовательность представляющих его символов (в данном случае имен операторов и операндов) случайна.

Итак, текст программы может рассматриваться как реализация случайных выборок имен операторов и операндов. Следовательно, процессу написания программы можно поставить в соответствие процесс выборки с возвратом (один и тот же оператор и операнд может появиться в выборке не один раз) из генеральной совокупности операторов и операндов до тех пор, пока последняя не будет исчерпана. Технически это можно представить так: имя каждого оператора и операнда записывается на отдельные «билеты», которые помещаются в урну и перемешиваются; затем вынимают наудачу один «билет», фиксируют имя оператора

или операнда и опускают его снова в урну, тщательно перемешивают ее содержимое; затем снова наудачу вынимают один «билет», снова фиксируют имя и т.д., повторяя все сначала. Этот процесс повторяется до тех пор, пока каждый оператор и операнд не появится в выборке хотя бы один раз.

Пусть η_1 – число операторов, а η_2 – число операндов. Величину $\eta = \eta_1 + \eta_2$ назовем словарем программы, она равна числу «билетов». Имя первого слова из словаря (оператора или операнда) входит в выборку при первом извлечении. Для того чтобы в выборку вошел следующий элемент, отличный от первого, может потребоваться некоторое число извлечений L_j .

Вообще, пусть L_r будет числом извлечений, следующих за выбором r –того билета до выбора нового $(r + 1)$ го билета включительно.

Тогда

$$L_r = 1 + L_1 + L_2 + \dots + L_{r-1}$$

Является объемом выборки на тот момент, когда в нее входит r – ый новый оператор или операнд.

Распределение случайной величины L_r совпадает с распределением первого успеха в последовательности испытаний Бернули.

Исходя из этого, за математическое ожидание длины программы с длиной словаря η приближенно можно принять следующее выражение:

$$M(L_\eta) = 0.9 \eta \log_2 \eta$$

или

$$L = 0.9 (\eta_1 + \eta_2) \log_2 (\eta_1 + \eta_2).$$

Последнее выражение и принимается за математическое ожидание длины программы, выраженное в количестве операторов и операндов.

Можно показать, что дисперсия

$$D(L_\eta) = (\pi^2 \eta^2) / 6$$

Тогда относительная ожидаемая погрешность будет

$$\delta = (\sqrt{D(L_\eta)}) / M(L_\eta) \approx 1 / (2 * \log_2 \eta)$$

Если один и тот же алгоритм поручить реализовать на каком-либо языке нескольким программистам, то написанные ими программы будут, конечно, варьировать по длине. Но, как следует из приведенных выше соотношений, а также из огромного числа статистических проверок, сделанных на основе уже написанных

программ, эти отклонения невелики. Практика показывает, что δ обычно не превосходит 10%.

Так как функция L - выпуклая, то имеет место неравенство Йенсена, исходя из которого можно получить еще одно приближенное выражение для длины программы:

$$L(\eta_1, \eta_2) = \eta_1 \log_2 \eta_1 + \eta_2 \log_2 \eta_2$$

Задание

1. Разработать программу для вероятностного моделирования процесса написания программы программистом с длиной словаря программы $\eta = 16, 32, 64, 128$.
2. С помощью разработанной программы получить статистические оценки:
 - длины программы L ,
 - дисперсии длины $D(L_\eta)$,
 - среднеквадратического отклонения ($\sqrt{D(L_\eta)}$),
 - относительной ожидаемой погрешности δ .
3. С помощью приведенных формул получить теоретические значения и сравнить их с результатами моделирования.
4. По тексту разработанной программы посчитать длину ее словаря и длину программы. Рассчитать длину программы по размеру ее словаря с помощью приведенных формул. Сравнить посчитанное по тексту значение длины текста программы, с длиной текста программы, полученной по формуле.
5. По первому и второму пунктам задания определить η_2^* – число единых по смыслу входных и выходных параметров представленных в сжатой без избыточной форме. Сравнить прогнозируемую длину программы с длиной программы, рассчитанной по тексту программы.

Содержание отчета

Отчет должен содержать:

- цель работы,
- текст программы,
- статистические и расчетные характеристики длин программ для заданных размеров словарей,
- рассчитанную и полученную по тексту длину разработанной программы.

Контрольные вопросы

1. Какие особенности процесса написания программы программистом учитываются в имитационной модели?
2. Что такое словарь программы?
3. Как связана длина программы с длиной словаря программы?
4. Как связана длина программы с числом операторов и операндов в программе?
5. Чему равна относительная ожидаемая погрешность длины программы?

Литература

1. Кайгородцев Г. И. Введение в курс метрической теории и метрологии программ : [учебник] / Г. И. Кайгородцев. - Новосибирск, 2009. - 191 с. : ил., табл.
2. Холстед М.Х. Начала науки о программах. - М.: Финансы и статистика, 1981.
3. Феллер В. Введение в теорию вероятностей и её приложения. Т.1. - М.: Мир, 1967.