

тами его словоформ), которое статистически неравномерно распределено по тексту в целом. (А как быть с видоизмененными терминами, поддающимися спряжению, склонению по падежам и т.д. – словоформами?). Проблема отождествления различных словоформ при построении частотного словаря может быть решена следующим способом: в каждом слове оставляется не более 5 литер (байтов), причем все гласные, кроме первой, выбрасываются. Этот прием опирается на традицию некоторых древних языков, который вообще обходится при записи без гласных. Подавляющая информативность именно согласных букв давно известна лингвистам. Об этом факте говорят и "нктр псхлг". Если в словах такого выжатого текста оставить еще и по одной гласной, то "чтен и понмн такг текст станв вобщ тривл". Значит, такое сжатие сохраняет смысл фразы и текста в целом.

Замечено также, что согласные образуют своеобразную решетку узлов, вокруг которой при помощи преимущественно гласных и окончаний, образно говоря, "колеблются" конкретные слова и словоформы: «ИнфОрмАтИкА», «нЕфОрмАльнО» (имеют схожий набор согласных букв «НФРМ...»). Учет одной гласной позволяет различить и эти слова: «Инфрм, Нефрм». Понятны ограничения, присущие данному формализованному подходу. Но он в основном позволил различить по смыслу группы слов и отождествить употребление слов, относящихся к одному понятию. Чисто количественные характеристики такого метода образования частотного словаря следует еще исследовать, так как в рассматриваемых текстах доли потерь и шума при построении словаря составляют порядка процента. Это выражается в формальной омонимии, например, при сжатии слов «ТРАНСлятор» и «ТРАНСпарант» и различении программой словоупотреблений «МОДуЛЬный» и «МОДуЛЬ». Метод построения словарей отбрасыванием псевдоокончаний пока не использовался из-за увеличения времени счета.

Именно используя понятие термина, открывается целый спектр всевозможных методов анализа компьютерного текста, в частности, статистических.

Современные пакеты программ достаточно просто обрабатывают введенные данные с использованием различных процедур, а после обработки уже можно выбрать метод, дающий наилучшие результаты.

Классификация статистических методов:

- по количеству анализируемых признаков (одномерные, двумерные, многофакторные);
- по статистическим принципам, лежащим в основе методов (параметрические и непараметрические);
- по зависимости или независимости сопоставленных выборок.

Литература

1. Колмогоров А.Н. Три подхода к определению понятия «количество информации» // Проблемы передачи информации. 1965. Т.1. №1, С. 3-11.
2. Яглом А.М., Яглом И.М. Вероятность и информация. М.: Наука, 1960.
3. Применение статистических методов анализа данных в научно-исследовательской работе: Учебно-методическое пособие / Авт.-сост.: А.И. Бочкин, Л.П. Колбасич. – Витебск: Издательство УО «ВГУ им. П.М. Машерова», 2006. – 39 с.
4. Пушкарев, А. В. О некоторых распределениях терминов в компьютерном тексте / А.В. Пушкарев // III Машеровские чтения. Математика. Информатика. Философия. Экономика. Юриспруденция: материалы республиканской научно-практической конференции студентов, аспирантов и молодых ученых, Витебск, 24-25 марта 2009 г. – Витебск: УО "ВГУ им. П.М. Машерова", 2009. – С. 78-79. – Библиогр.: с. 79 (2 назв.).

ИМИТАЦИОННОЕ МОДЕЛИРОВАНИЕ СХЕМ КОРРЕКЦИИ РАВНОВЕРОЯТНОСТНЫХ БЛОКОВ

М.А. Чиркин

Область применения стохастических устройств в технике постоянно увеличивается, а их сложность возрастает. В связи с этим возникает необходимость разработки систем автоматизации моделирования и исследования стохастических устройств [1].

Одними из важнейших компонентов стохастических устройств являются равновероятностные блоки и схемы их коррекции, которые предназначены для формирования последовательностей равновероятностных случайных чисел. Они служат:

- для имитации часто встречающихся равномерно распределенных случайных воздействий;
- в качестве входных параметров при формировании потоков случайных величин, подчиняющихся произвольным распределениям.

Как правило, равновероятностные блоки реализуются в виде электронных устройств. Для повышения качества генерируемых последовательностей чисел могут использоваться различные схемы коррекции. Критериями качества равновероятностных чисел являются степень близости вероятности их появления к теоретической и отсутствие корреляционных связей.

Различные способы построения схем коррекции показаны на рисунке 1.

Для примера рассмотрим имитационное моделирование работы устройства коррекции на основе способа Неймана-Точера [2]. Этот способ был разработан применительно к программной реализации на ЭВМ. Его физическим аналогом является датчик Полака. Сущность способа Неймана-Точера заключается в следующем: входная последовательность $\dots, x_i, x_{i+1}, \dots$ разбивается на пары $\dots, (x_{i-2}, x_{i-1}), (x_i, x_{i+1}), (x_{i+2}, x_{i+3}), \dots$. Каждой паре ставится в соответствие некоторое промежуточное число η_{2i} . Если $x_i = x_{i+1}$, то есть $\eta_{2i} = 00$ или $\eta_{2i} = 11$, то это соответствует выработке холостого числа. Если $x_i \neq x_{i+1}$, то η_{2i} ставится в соответствие число z_j , например, по следующему правилу:

$$z_j = \begin{cases} 1, & \text{если } \eta_{2i} = 01 \\ 0, & \text{если } \eta_{2i} = 10 \end{cases}$$

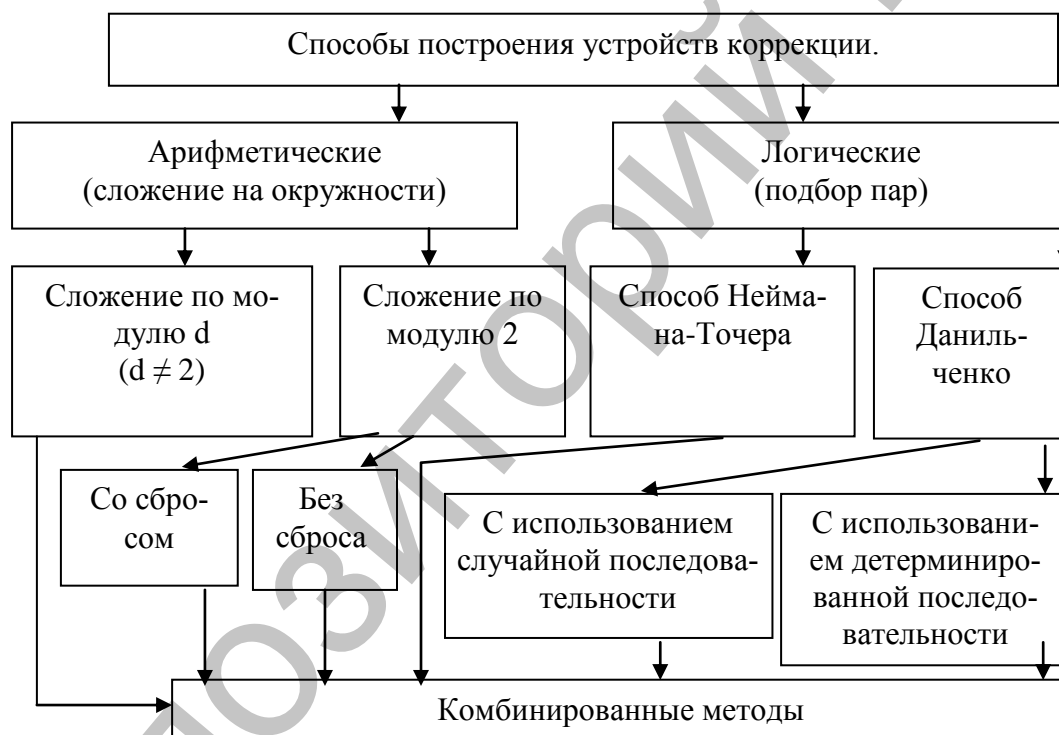


Рис. 1. Классификация способов построения устройств коррекции.

Схема устройства коррекции, основанного на способе Неймана-Точера приведена на рис. 2.

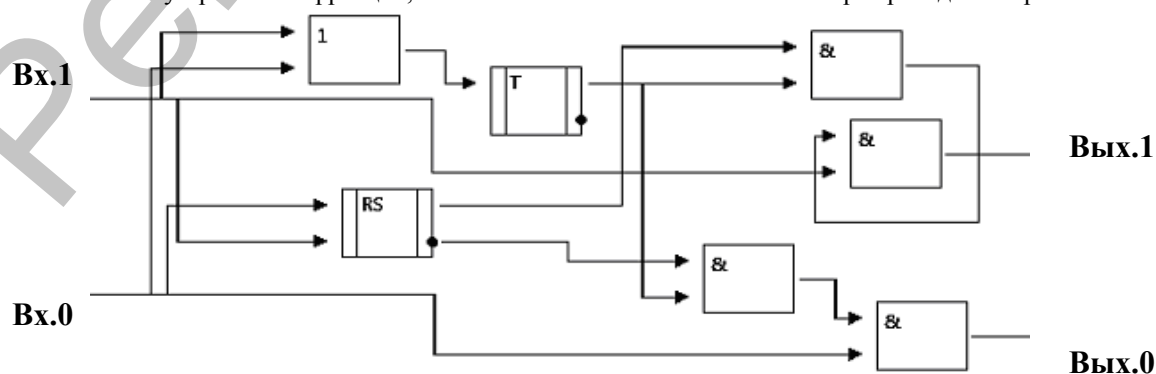


Рис. 2. Функциональная схем устройства коррекции, основанного на способе Неймана-Точера.

Для моделирования устройств коррекции было разработано приложение, включающее:

- визуальный редактор схем электронных устройств;
- библиотеку для моделирования компонентов электронных устройств;
- набор генераторов случайных сигналов, эмулирующих равновероятностные блоки;
- набор средств анализа выходных последовательностей;
- набор средств визуального анализа состояния моделируемого электронного устройства.

Одним из результатов исследования имитационной модели является временная диаграмма работы устройства (рис.3).

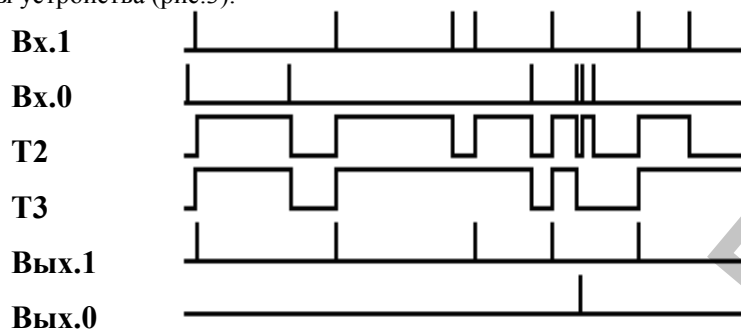


Рис. 3. Временная диаграмма работы устройства коррекции, основанного на способе Неймана-Точера.

Таким образом, было разработано программное средство, позволяющее на стадии проектирования имитировать работу устройства коррекции равновероятностных блоков. Это приложение позволяет выполнять широкий круг экспериментов над моделями таких устройств и отличается от существующих аналогов тем, что оно учитывает как математические аспекты моделирования, позволяющие задавать характеристики выходных последовательностей равновероятностных блоков, так и электронную схему моделируемого устройства.

Литература

1. Четвериков В.Н., Баканович Э.А., Меньков А.В. Вычислительная техника для статистического моделирования. М., 1978.
2. Четвериков В.Н., Баканович Э.А. Стохастические вычислительные устройства систем моделирования. М., 1989