



КЛАССИФИКАЦИЯ СИГНАЛОВ СТОХАСТИЧЕСКОГО АВТОГЕНЕРАТОРА МЕТОДАМИ РАСПОЗНАВАНИЯ ОБРАЗОВ

Д.Ю.Васин, В.П.Громов

В работе исследуется новая технология оценки состояния нелинейной динамической системы, в которой может наблюдаться хаос. Технология базируется на последовательном использовании методов распознавания образов. На основе предложенной технологии синтезирована система классификации сигналов стохастического генератора (СГ) с 1.5 степенями свободы. Предлагаемый подход к оценке состояний СГ позволяет реализовать эффективное решение следующих задач обработки сигналов СГ:

- обоснованный выбор параметров и алгоритмов амплитудно–временного квантования и сегментации сигналов;
- построение систем компактного представления фрагментов сигналов;
- формирование машинных банков обучающих данных;
- построение систем решающих правил классификации сигналов;
- синтез систем классификации состояний СГ.

Открытие динамического хаоса [1,2] ознаменовало начало принципиально нового этапа в развитии нелинейной динамики. Для этого этапа характерно накопление теоретических и экспериментальных знаний о генерации сигналов СГ, что интересно не только для формирования собственно теории динамического хаоса, но и для возможных практических приложений, связанных с кодированием, передачей по каналам связи и декодированием информационных сообщений с повышенными требованиями к их конфиденциальности, а также в задачах оценки состояния технических и биологических систем по их сигнальным характеристикам.

Для описания хаотических сигналов в настоящее время часто пользуются понятием фрактальной размерности [2]. В реальных условиях оценка фрактальной размерности производится методами выборочной статистики на основе достаточно длительных цифровых реализаций сигнала с выполнением большого объема вычислений.

В данной работе исследовалась возможность получения эффективных алгоритмов описания сигналов СГ методами распознавания образов (РО). Перспективность использования методов РО в задачах описания сигналов СГ обусловлена, во–первых, наличием мощных методов сжатия исходных экспериментальных данных с сохранением релевантной, по отношению к целевой задаче, информации и, во–вторых, большой популярностью РО в практических приложениях обработки сигналов [3–8].

В качестве СГ использовался автогенератор радиодиапазона с 1.5 степенями свободы, функциональная блок–схема которого приведена на рис.1.

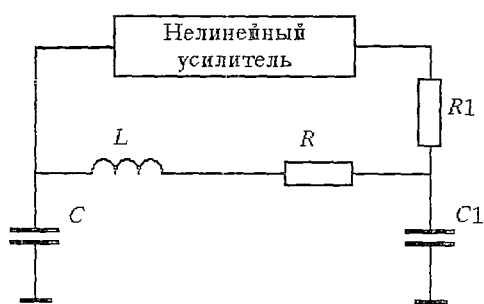


Рис. 1. Функциональная блок-схема генератора, используемого в работе

Данный генератор представляет собой вариант генератора кольцевого типа с 1.5 степенями свободы [9,10]. Генератор реализует фейгенбаумовский сценарий перехода к хаосу [1,2,9,10]. Сигналы для идентификации отбирались на основании перемен, происходящих в проекции на плоскости фазового портрета системы, при изменении ее внешних параметров. Всего было отобрано 11 различных типов (классов) сигналов. Качественный вид фазовых портретов системы для каждого типа сигнала приведен на рис. 2 (цифра справа от рисунка указывает на номер соответствующего типа).

Задачу идентификации сигналов СГ определим следующим образом.

Пусть на выходе устройства, реализующего физическую модель СГ, наблюдается на промежутке времени $0 < t < T$ сигнал $x(t)$ со следующими свойствами:

- 1) сигнал $x(t)$ ограничен, то есть $x(t) < A$, $-\infty < t < \infty$;
- 2) сигнал $x(t)$ абсолютно интегрируем на интервале $-\infty < t < \infty$;
- 3) сигнал $x(t)$ имеет конечное число максимумов, минимумов и точек разрыва;

4) для сигнала $x(t)$ существует его Фурье – образ $X(\nu)$.

По массивам обучающих реализаций сигнала необходимо синтезировать структуру распознающего устройства, позволяющего с заданной степенью надежности правильно устанавливать принадлежность текущей реализации сигнала $x(t)$ к одному из заданных на этапе обучения классов и тем самым идентифицировать состояние СГ.

Эксперименты по распознаванию проводились на базе АСНИ «МИСС» [11], разработанной в Научно-исследовательском институте прикладной математики и кибернетики (НИИ ПМК), и устройства, реализующего аппаратно вышеописанную модель СГ, сконструированного на кафедре теории колебаний и автоматического регулирования ННГУ.

Решение сформулированной выше задачи осуществлялось по схеме, характерной при построении системы распознавания образов [12].

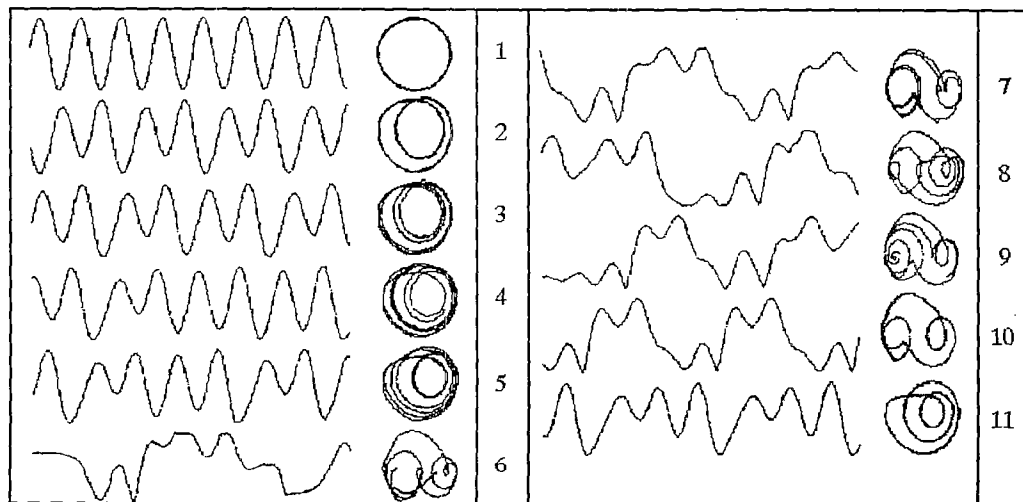


Рис. 2. Осциллограммы сигналов, получаемых с генератора, использованного в работе, и фазовые портреты, соответствующие данной реализации сигнала

1. Осуществление обоснованного выбора параметров квантования и сегментации сигналов СГ методом спектрального и корреляционного анализа [12–14].

2. Выбор и реализация эффективного алгоритма сжатия исходных цифровых данных сигналов СГ на основе метода построения «хорошо приспособленного» базиса [3].

3. Формирование машинного банка обучающих данных, включающего сжатое цифровое описание 1650 фрагментов сигналов СГ с различными фазовыми портретами (см. рис. 2).

4. Анализ дискриминантных свойств цифровых признаков сигналов СГ методами РО и построение на его основе системы решающих правил и классифицирующих признаков.

5. Синтез распознающего устройства на базе полученных решающих правил и дискриминантных признаков для классификации сигналов СГ на указанные выше 11 классов (см. рис. 2).

6. Осуществление оценки по машинному банку обучающих данных устойчивости и достоверности функционирования синтезированного устройства распознавания сигналов СГ.

Для реализации данной технологии был разработан комплекс программ по накоплению, предварительной обработке и распознаванию сигналов СГ. В целях минимизации ошибок формирования машинного банка обучающих данных, отбор сигналов в соответствии с приведенной выше классификацией и контроль правильности сегментации велись в интерактивном режиме. В дальнейшем, на этапах сокращения избыточности описания сигналов, построения дискриминантных функций и непосредственно на этапе распознавания комплекс работал автономно, без участия оператора.

Остановимся на практических вопросах реализации некоторых пунктов описанного выше общего плана решения поставленной задачи. Для выбора оптимальной частоты квантования сигналов СГ был вычислен спектр Фурье для каждого класса сигналов. Спектр мощности сигнала определялся по методу быстрого преобразования Фурье (БПФ) [13, 14]. Для уменьшения эффектов, связанных с конечностью выборки, применялась оконная функция Хэмминга [13]. По результатам анализа спектрограмм для всех типов классифицируемых сигналов с использованием теоремы отсчетов и сохранением дискриминантных свойств сигналов СГ для надежной их идентификации, были выбраны следующие значения оптимальных параметров оцифровки сигналов: частота квантования по времени – 10 КГц; число уровней квантования по амплитуде – 1024.

Корреляционный анализ и исследование амплитудно–временных свойств сигналов СГ позволило синтезировать алгоритм сегментации сигналов СГ в следующем виде.

1. Фиксация локального максимума сигнала по его текущей цифровой реализации длительностью 100 мс.

2. Выбор в качестве представительного фрагмента сигнала отсчетов текущей реализации на временном интервале ± 6 мс относительно зафиксированного локального экстремума.

Данный алгоритм позволяет максимально совместить начальные фазы выбираемых фрагментов сигналов СГ и описать их одинаковым числом отсчетов $N = 120$, что является существенно важным моментом для последующих этапов кодирования и распознавания сигналов СГ.

На следующем этапе был сформирован машинный банк обучающих фрагментов сигналов СГ для 11 выбранных классов. Общий объем банка данных составил

$$N_0 = K_{\text{кл}} N_{\text{кл}},$$

где $K_{\text{кл}} = 11$ – число классов обучения; $N_{\text{кл}} = 150$ – число обучающих фрагментов заданного класса.

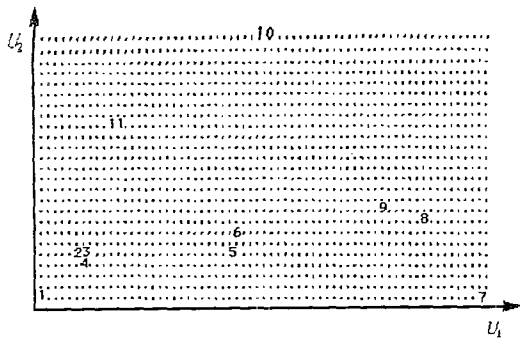


Рис. 3. Взаимное расположение эталонных сигналов СГ в двумерном «хорошо разделяющем» подпространстве $R_{U_1 U_2}$

проведения эксперимента. В данной работе она решалась по методу «хорошо приспособленного» базиса [3], как наиболее эффективного метода сжатия данных при заданной допустимой ошибке кодирования.

В результате проведения данного эксперимента было сокращено описание объектов со 120 до 15 отсчетов, то есть коэффициент сжатия информации равен 8. Это означает, что по оставшимся 15 информативным отсчетам можно с заданной степенью точности E восстановить исходный сигнал, то есть 120 отсчетов.

Для качественного представления и грубой оценки дискриминантных свойств построенного 15-мерного пространства признаков сигналов СГ были использованы известные методы многомерного шкалирования [7] и кластерного анализа [4,8], то есть представления многопризнаковых объектов в двумерном пространстве и выявления скоплений (таксонов) объектов. На рис. 3 представлено взаимное расположение эталонных сигналов СГ в двумерном «хорошо разделяющем» подпространстве $R_{U_1 U_2}$. За эталон принимался средний объект класса. Здесь U_1, U_2 – собственные вектора взвешенной матрицы ковариации типа

$$D = [d_{ij}]_{i=1, n; j=1, n},$$

где n – размерность пространства признаков; $d_{ij} = \frac{1}{N_1 N_2} \sum_{k=1}^{N_1} \sum_{l=1}^{N_2} [(x_i^k - y_l^j)(x_j^k - y_l^j)] r_{k,l}^a$,

$r_{k,l} = [\sum_{i=1}^n (x_i^k - y_l^j)^2]^{1/2}$, $a=1$. В результате проведенных экспериментов по многомерному шкалированию было установлено, что сигналы из классов 1–4 (условно «детерминизм») располагаются в пространстве признаков на достаточном расстоянии от сигналов из классов 5–9 (условно «хаос»), а те, в свою очередь, обособлены от сигналов из классов 10,11 (запредельные состояния) (см. рис. 3). Однако, плотность точек внутри этих трех групп достаточно однородна, что не позволило разделить сигналы на классы внутри трех полученных скоплений методами кластерного анализа. К тому же, несмотря на оптимизацию алгоритма кластеризации по времени, он все равно достаточно трудоемок в работе.

Для решения задачи разделения объектов на 11 классов, то есть для формирования и выбора оптимальной совокупности решающих правил классификации сигналов СГ, нами использовались методы: эталона, k - ближайших соседей [4], а также методы распознавания на основе решающих правил в форме пороговых логических функций [5].

Так называемые непараметрические методы распознавания образов (эталона, k - ближайших соседей) являются асимптотическими, то есть обеспечивают достаточную точность оценивания только лишь при неограниченном увеличении объема обучающих выборок. В случае использования метода эталона за эталон выбирался некий средний объект класса, то есть, вообще говоря, несуществующий реально объект. Средний объект класса определялся как

Цифровое представление и сегментация исходных обучающих сигналов производилась в соответствии с указанными выше параметрами квантования и сегментации.

Одной из важнейших задач, которую приходится решать в процессе построения любой системы распознавания, является задача сокращения избыточности исходного описания, то есть задача кодирования. Как говорилось ранее, рассмотренные сигналы дискретизированы с достаточной степенью избыточности. Поэтому задаче кодирования отводится одно из важнейших мест во всей цепочке

$$\tilde{x}_i = \frac{\sum_{j=1}^M x_i^j}{M},$$

где $j = 1, 2, 3, \dots, M$, \tilde{x}_i – i -й отсчет среднего объекта класса; x_i^j – i -й отсчет j -го объекта класса.

В [4] приводятся оценки вероятности возникновения ошибок при использовании правила k -ближайших соседей. Однако, практическим препятствием для рекомендации данных методов в качестве основных для синтеза на их основе системы идентификации сигналов СГ является то обстоятельство, что для достижения теоретических оценок вероятности возникновения ошибки, необходимо сохранять в памяти большое число образов, о которых известна принадлежность их некоторому классу. Кроме того, при осуществлении классификации необходимо вычислять расстояния между каждым классифицируемым образом и всеми образами, хранящимися в памяти системы. При больших объемах обучающих выборок это обстоятельство может вызвать серьезные вычислительные трудности. При произвольном (в том числе малом) размере выборки достоверность распознавания становится недопустимо низкой.

Более привлекательными поэтому представляются алгоритмы распознавания на базе построения разделяющих функций в форме пороговых логических функций. Суть данного метода состоит в том, что по совокупности исходных данных строится функция классификации F в форме

$$F = \bigcup_p \prod_q (x_{ip,q} > a_{p,q})^{\sigma_{p,q}}, \quad (*)$$

то есть в дизъюнктивно–нормальной форме от простых высказываний $(x_{ip,q} > a_{p,q})^{\sigma_{p,q}}$, состоящих в проверке истинности сравнений некоторых исходных признаков $x_{ip,q}$ пороговыми значениями $a_{p,q}$, причем высказывание входит в формулу (*) без отрицания, если $\sigma_{p,q} = 0$, и с отрицанием, если $\sigma_{p,q} = 1$.

Предполагается также, что $F = 1$ на множестве обучающих объектов x , а поиск (*) осуществляется в пространстве признаков \mathbf{R}^n .

Вид функции F предполагает аппроксимацию истинной разделяющей поверхности гиперплоскостями, направляющие орты которых коллинеарны отдельным ортам исходного пространства признаков. Такой способ приближения истинных разделяющих поверхностей допустим для ряда задач классификации без существенных потерь в качестве распознавания при значительном сокращении числа параметров и необходимого объема арифметических операций.

По полученным выше алгоритмам, параметрам и процедурам дискретизации, сегментации, кодирования и классификации была синтезирована система автоматического распознавания сигналов СГ, позволяющая по поступающим на ее вход аналоговым сигналам автоматически определять их принадлежность к одному из 11 типов состояния СГ. Контрольная проверка качества распознавания этой системы по независимым входным сигналам СГ показала, что уровень ошибок классификации при функционировании системы в реальных условиях не превышает зафиксированный на этапе обучения, а время узнавания сигнала СГ не превышает 1 с.

Основные результаты классификации приведены в таблице. Как видно из

Таблица

Методы	Эталона	1 ближайшего соседа	10 ближайших соседей	50 ближайших соседей	Логических функций
Интегральный процент ошибок распознавания	33.94	5.21	8.73	18.79	6

таблицы, хуже всего работает классификатор на основе метода эталона, хотя он выигрывает по сравнению с другими методами благодаря простоте реализации вычислений.

В результате проведения данной работы было выяснено, что задачу идентификации сигналов СГ можно с достаточной степенью надежности решить, применяя для этого классические методы и приемы теории РО, что позволяет в несколько раз сократить время решения этой задачи по сравнению с уже существующими способами идентификации сигналов СГ.

В заключение авторы выражают глубокую благодарность В.Д. Шалфееву и Н.Ф. Рулькову за помощь и поддержку при проведении данной работы.

Работа выполнена при финансовой поддержке Российского фонда фундаментальных исследований (проект 93-02-15424).

Библиографический список

1. Рабинович М.И., Трубецков Д.И. Введение в теорию колебаний и волн. М.: Наука, 1984.
2. Шустер Г. Детерминированный хаос. Введение. М.: Мир, 1988.
3. Васин Ю.Г. «Хорошо приспособленные» базисы и задачи обработки экспериментальной информации. Учебное пособие. Горький : Изд-во Горьк. ун-та, 1979.
4. Ту Дж., Гонсалес Р. Принципы распознавания образов / Под ред. Ю.И. Журавлева. М.: Мир, 1978.
5. Распознавание образов и медицинская диагностика / Под ред. Ю.И. Неймарка. М.: Наука, 1972.
6. Горелик А.Л., Скрипкин В.А. Методы распознавания. М.: Высшая школа, 1977.
7. Александров В.В., Горский Н.Д. Алгоритмы и программы структурного метода обработки данных. Л.: Наука, 1983.
8. Дюран Б., Оддел П. Кластерный анализ. М.: Статистика, 1977.
9. Дмитриев А.С., Кислов В.Я. Стохастические колебания в автогенераторе с инерционным запаздыванием первого порядка // Радиотехника и электроника. 1984. Т.29, № 12. С. 2389.
10. Волковский А.Р., Рульков Н.Ф. Применение одномерных отображений для экспериментального исследования стохастической динамики автогенератора // Письма в ЖТФ. 1988. Т.14, вып.16. С. 1508.
11. Исследование и разработка методов, математических моделей описания, хранения и принятия решений, принципов построения адаптивных автоматизированных систем обработки сложной графической информации, экспериментальных данных и изображений // Отчет НИИ ПМК № гос. регистрации 0186.0123541. Горький, 1990.
12. Алиев Т.А. Экспериментальный анализ. М.: Машиностроение, 1991.
13. Гольденберг Л.М., Матюшкин Б.Д., Поляк М.Н. Цифровая обработка сигналов. Справочник. М.: Радио и связь, 1985.
14. Отнес Р., Энксон Л. Прикладной анализ временных рядов. Основные методы. М.: Мир, 1982.

*НИИ прикладной математики
и кибернетики*

*Поступила в редакцию 29.04.94
после переработки 12.07.94*

CLASSIFICATION OF STOCHASTIC AUTOGENERATOR SIGNALS BY PATTERN RECOGNITION METHODS

D. Yu. Vasin, V.P. Gromov

In this research a new technology for estimating a state of the nonlinear dynamical system in which chaos can be observed has been studied. The technology lies upon sequential employment of pattern recognition methods. On the basis of this technology a system of classifying stochastic generator (SG) signals with 1.5 degrees of freedom has been synthesized.

The suggested approach for stochastic generator state estimation makes possible to solve efficiently the following problems of stochastic generator signal processing :

- choosing on a substantial basis parameters and algorithms for amplitude-time quantification and for segmenting SG signals ;
- designing the systems for compact representation of SG signal fragments ;
- designing the PC-based banks of training data ;
- deriving the systems of decision rules for SG signal classification;
- synthesizing the systems for classifying SG states.



Васин Дмитрий Юрьевич – родился в 1970 году в г. Горьком. В 1992 году окончил радиофизический факультет Нижегородского государственного университета. После его окончания работает в НИИ ПМК. Область научных интересов – теория распознавания образов, машинная графика и системы компьютерной обработки изображений.



Громов Владимир Петрович – родился в 1941 году. В 1968 году окончил физический факультет Горьковского государственного университета. С 1968 года работает в НИИ ПМК – заведующий сектором медицинской информатики. Область научных интересов – компьютерная обработка экспериментальных данных, распознавание образов, инженерия знаний.