Из вышесказанного следует критерий выбора оптимального алгоритма распознавания a^* из совокупности проверяемых алгоритмов $a \in \Omega_A$ с одновременной оценкой их параметров:

$$a^* = \arg \max_{a \in \Omega_A} S(a)$$

при дополнительном требовании к поведению характеристики распознавания (монотонная невозрастающая функция) для алгоритма a^* . При ограниченном объеме обучающих данных мы имеем дело с оценками параметров анализируемых алгоритмов принятия решения. Поэтому требует внимания выбор информативного состава признаков. Введем следующее определение информативности совокупности признаков распознавания (ПР) размерности L по отношению к совокупности ПР размерности $L_1 < L$, при условии, что последняя содержится в первой и выбран байесов алгоритм принятия решения: $I(L | L_1) = S(L) - S(L_1)$. Введенное определение корректно, поскольку значение эффективности распознавания при отсутствии признаков на входе алгоритма классификации равно априорной вероятности появления объектов заданного класса.

Вопросы реализации алгоритма по эмпирическим данным ограниченного объема с использованием методов многомерного статистического анализа [4] рассмотрены в [3] и в данной работе не затрагиваются.

Заключение.

В работе рассмотрена задача поиска наилучшего классификатора и используемого состава признаков распознавания из числа доступных алгоритмов принятия решения в условиях ограниченного состава обучающей выборки и неполного вероятностного описания содержащейся в ней информации.

В рамках статистической проверки двувальтернативных гипотез предложены показатель качества и характеристика распознавания, использование которых позволяет найти оптимальный классификатор при наличии соответствующей совокупности алгоритмов формирования решения и процедуру его обучения по доступной обучающей выборке.

1. Тартаковский Г.П. Теория информационных систем. М.: Физматкнига, 2005.
2. Рындин Ю.Г., Тартаковский Г.П., Тюрин В.С. //Радиотехника и электроника, 2004, Т.49, вып. №9, стр. 1066.
3. Прохоров М.Е. Оценка качества распознавания при наличии единственной обучающей выборки ограниченного объема. //Радиотехника и электроника, 2007, Т.52, вып. №1, стр. 28-32.
4. Эфрон Б. Нетрадиционные методы многомерного статистического анализа. М.: Финансы и статистика, 1988.

Шульгин П.В., Муллин В.В., Ханмагомедов Р.Х.

Оценка параметров кодирования при использовании алгоритма Шеннона-Фано

*ФГКВОВУ ВО «Пермский военный институт войск национальной гвардии
Российской Федерации»
(Россия, Пермь)*

doi: 10.18411/sr-10-12-2020-33

idsp: sciencerussia-10-12-2020-33

Аннотация

Статья посвящена оценке параметров кодирования при использовании алгоритма Шеннона-Фано. Проведенные численные исследования показали, что применение данного алгоритма кодирования способствует уменьшению объема передаваемого сообщения, что позволяет сократить время передачи отправляемого сообщения.

Ключевые слова: Шеннон, кодирование, алгоритм, алфавит, вероятность, численное исследование.

Abstract

The article is devoted to the estimation of encoding parameters using the Shannon-Fano algorithm. Numerical studies have shown that the use of this encoding algorithm reduces the volume of the transmitted message, which reduces the transmission time of the sent message.

Keywords: Shannon, coding, algorithm, alphabet, probability, numerical investigation.

Теория информации - наука о свойствах кодов и их пригодности для достижения поставленной цели. Кодирование информации — это процесс её преобразования из формы, удобной для непосредственного использования, в форму, удобную для передачи, хранения, автоматической переработки и сохранения от несанкционированного доступа. Различают три вида кодирования информации: криптографическое, помехоустойчивое и эффективное. Все перечисленные виды кодирования преобразовывают представление информации.

Современные системы связи работают на основе передачи сообщений в цифровом виде. Процесс передачи от источника к приемнику нередко сопровождается воздействиями на информацию негативных факторов, которые непосредственно искажают передаваемую информацию, в следствии чего появляются ошибки или передаваемая информация вовсе теряется.

В связи с этим были созданы эффективные методы контроля точности и полноты передаваемых сообщений, иначе называемых методами помехоустойчивого кодирования, в которых реализуется следующий последовательный алгоритм:

- 1) составление алфавита сообщения;
- 2) кодирование алфавита с помощью двоичного кода;
- 3) кодирование сообщения для передачи;
- 4) декодирование сообщения.

В данной работе будет рассматривается кодирование сообщения с помощью алгоритма Шеннона-Фано, который можно представит в виде следующей последовательности действий:

- 1) составление алгоритма передаваемого сообщения;
- 2) определение вероятностей появления символов;
- 3) сортировка символов первичного алфавита в порядке убывания вероятностей их появления;
- 4) разделение символов полученного алфавита на две части, суммарные вероятности символов которых максимально близки друг другу;
- 5) продолжение процесса объединения символов согласно пунктов 3 и 4 до тех пор, пока в новом алфавите не останется двух символов;
- 6) кодирование символов нового алфавита с помощью двоичного кода, при этом в префиксном коде для первой части алфавита присваивается «1», а второй части «0»;
- 7) рекурсивное разделение полученных частей, которые заменяются на соответствующие двоичные цифры в префиксном коде
- 8) процесс создания двоичных комбинаций продолжается до тех пор пока не будет закодирован исходный алфавит.

На рис. 1 представлен пример кодирования сообщения «военное положение» с помощью алгоритма Шеннона-Фано.

Рисунок 1. Представление алгоритма Шеннона-Фано.

В табл. 1. представлен алфавит сообщения «военное положение».

Таблица 1

Алфавит сообщения

№	Символ алфавита	Количество символов	Вероятность появления символа
2	о	4	0.250
3	е	4	0.250
4	н	3	0.188
1	в	1	0.063
5	п	1	0.063
6	л	1	0.063
7	ж	1	0.063
8	и	1	0.063

В табл. 2. представлена двоичная кодировка алфавита сообщения «военное положение».

Таблица 2

Кодировка алфавита

№	Символ алфавита	Двоичная кодировка символа
1	о	11
2	е	10
3	н	011
4	в	010
5	п	0011
6	л	0010
7	ж	0001
8	и	0000

В табл. 3 представлены результаты кодирования сообщения «военное положение» с помощью алгоритма Шеннона-Фано и равномерного кода.

Таблица 3

Кодирование сообщения

Вид кодирования	Алгоритм Шеннона-Фано	Равномерное кодирование
Результат кодирования	010 11 10 011 011 11 10 0011 11 0010 11 0001 10 011 0000 10	000 001 010 011 001 010 100 001 001 110 010 011 111 010
Количество информации в сообщении, бит	45	48
Количество информации на каждый символ, бит	2,81	3

Из табл. 3 видно, что алгоритм Шеннона-Фано дает меньшую избыточность кода по сравнению с равномерным кодированием.

Алгоритм Шеннона-Фано был реализован в виде программы на языке Free Pascal [2].

Ниже представлены результаты численных исследований параметров помехоустойчивого кодирования сообщений с помощью алгоритма Шеннона-Фано.

Рисунок 2. Среднее количество бит на символ (ось x - количество символов, ось y - среднее количество бит на символ).

Из графика, представленного на рис. 2, видно, что при увеличении количества символов в исходном сообщении среднее количество бит на символ имеет тенденцию к уменьшению, показывающую, что с ростом длины сообщений избыточность передаваемых сообщений будет снижаться.

Рисунок 3. Распределение вероятностей появления символа (ось x - вероятность появления символа в сообщении, y - среднее количество бит на символ).

Из графика, представленного на рис. 3, видно, что при увеличении вероятности появления символа среднее количество бит на символ уменьшается. То есть наиболее часто встречаемый в сообщении символ кодируется наименьшим количеством бит, что и позволяет уменьшить длину передаваемого сообщения, снизив его избыточность.

Рисунок 4. Изменение среднего количества бит на символ в зависимости от длины исходного сообщения (ось x — длина исходного сообщения, ось y — отношение среднего количества бит на символ при использовании алгоритма Шеннона-Фано к количеству бит на символ при равномерном кодировании; 1 — семь символов в сообщении, 2 — восемь символов в сообщении, 3 — девять символов в сообщении, 4 — десять символов в сообщении).

Из графика, представленного на рис. 4, видно, что использование алгоритма Шеннона-Фано дает возможность значительно уменьшить объем передаваемого закодированного сообщения. Направленность изменения среднего количества бит на символ, полученных при различной длине исходного сообщения, показывает эффективность кодирования с помощью алгоритма Шеннона-Фано.

Таким образом, кодирование сообщений с помощью алгоритма Шеннона-Фано сокращает длину передаваемого сообщения, уменьшая объем передаваемой информации, а, следовательно, повышает надежность и оптимизирует время передачи сообщения.

1. Золотарёв В.В., Овечкин Г.В. Основы современных технологий помехоустойчивого кодирования. История. Возможности. Перспективы. Справочник / Под. ред. чл.-кор. РАН Ю.Б. Зубарева. — М.: Горячая линия-Телеком, 2004.
2. Шульгин П.В., Греков А.В., Старенков И.А. Основы алгоритмизации и программирования: учебное пособие — Пермь: Изд-во ПВИ войск национальной гвардии, 2017. — 125 с.

Ярошенко Е.А., Дубинин Г.В.

**Разработка алгоритма работы автоматизированной системы управления
испытаниями автомобильных шин**

*Донской государственный технический университет
(Россия, Ростов-на-Дону)*

doi: 10.18411/sr-10-12-2020-34

idsp: sciencerussia-10-12-2020-34

Аннотация

В статье описывается алгоритм автоматизированной системы, обрабатывающая визуальную информацию для динамической оценки состояния качества резинотехнических изделий в процессе их старения.