

2. Ивахненко А. Г. Метод группового учета аргументов в задачах прогнозирования. — К.: Ин-т кибернетики АН УССР, 1976.—25 с.
3. Ивахненко А. Г., Мюллер Й. А. Самоорганизация прогнозирующих моделей. — К.: Техника, 1985.—223 с.
4. Малайчук В. П., Мозговой А. В. Обработка информации в средствах и системах неразрушающего контроля. — Д.: Изд-во ДГУ, 1992.—168 с.
5. Малайчук В. П., Мозговой А. В., Петренко А. Н. Математический аппарат решения задач неразрушающего контроля // Матер. десятой юбилейной междунар. конф. и выставки «Современные методы и средства неразрушающего контроля и технической диагностики». — К.: УИЦ «Наука. Техника. Технологія», 2002.—С. 76—87.

**THE DERIVATION OF DECISION RULES OF NON-DESTRUCTIVE TESTING ON THE BASIS OF THE METHOD OF THE GROUP REGISTRATION OF ARGUMENTS**

N. Lysenko

The efficiency of the empirical decision rules constructed on the basis of the method of the group registration of arguments is investigated. The comparison of the empirical and optimum decision rules is carried out by computing experiments. The estimates for the probabilities of the identification are derived and their relation to the length of the initial sampling is investigated. The problem of the division of the sampling into learning and checking ones is analysed.

УДК 004.056(075.8)

© Е. В. Дзыгин

Дніпропетровський національний університет

**АДАПТИВНЫЙ АЛГОРИТМ СЖАТИЯ  
НЕСТАЦИОНАРНЫХ ЗНАКОВЫХ СООБЩЕНИЙ**

Розроблено адаптивний алгоритм стиснення нестационарних знакових повідомлень. Експериментально досліджено зміну коефіцієнта стиснення у залежності від довжини і типу повідомлення. Розроблено адаптивний алгоритм стиснення. Отримано практичні рекомендації з виявлення змін статистики стиснуваного повідомлення. Проведено порівняння розробленого алгоритму стиску з базовими алгоритмами.

В настоящее время есть множество алгоритмов сжатия, которые хорошо зарекомендовали себя при сжатии стационарных знаковых сообщений. В случае, если статистика сообщения нестационарна, то степень сжатия значительно уменьшается. Эта проблема актуальна ввиду появившейся в современных компьютерных системах тенденции к комбинированию в одном файле различных типов данных (тексты на различных языках, аудиоданные, графическая информация и т. п.), служащих для решения одной задачи. Здесь предлагается адаптивный алгоритм сжатия, определяющий изменения в статистике сообщения с помощью отслеживания изменения коэффициента сжатия.

В качестве базовых алгоритмов используется адаптивный алгоритм Хаффмана и алгоритм LZW. Алгоритм Хаффмана является статистическим, а LZW — словарным. В статистическом сжатии каждому символу присваивается код, основанный на вероятности его появления в тексте. Высоковероятные символы получают короткие коды, и наоборот. В словарном методе группы последовательных символов или «фраз» заменяются кодом. Замененная фраза может быть найдена в некотором «словаре».

Сжатие может осуществляться по априорным

данным о статистических свойствах исходных сообщений, или же проводится по технологии индивидуального кодирования каждого сообщения. Адаптивный алгоритм Хаффмана и алгоритм LZW реализуют индивидуальное кодирование для каждого сообщения. В процессе своей работы они приспосабливаются к статистике сообщения, и качество сжатия улучшается.

Идея написания адаптивного алгоритма сжатия возникла после проведения ряда исследований изменения коэффициента сжатия от длины прочитанного сообщения. Коэффициент сжатия определяется как отношение длины исходного сообщения к длине сжатого сообщения:

$$K = N/lL,$$

где  $N$  — количество знаков в исходном сообщении,  $l$  — длина одного знака сообщения,  $L$  — длина сжатого сообщения.

На рис. 1 представлен график изменения коэффициента сжатия от длины прочитанного сообщения. Для сжатия применялся текст на английском языке. Длина сообщения 1700 знаков. Видно, что на начальном этапе сжатия использовать алгоритм LZW невыгодно, поскольку приблизительно до 500

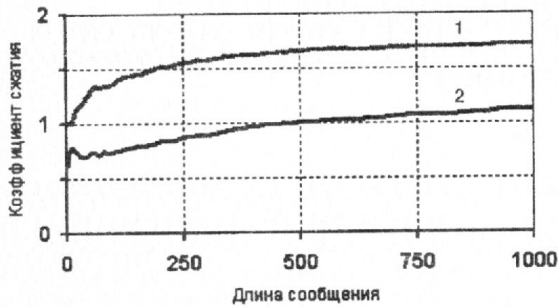


Рис. 1. Изменение коэффициента сжатия от длины прочитанного сообщения для алгоритмов Хаффмана (1) и LZW (2)

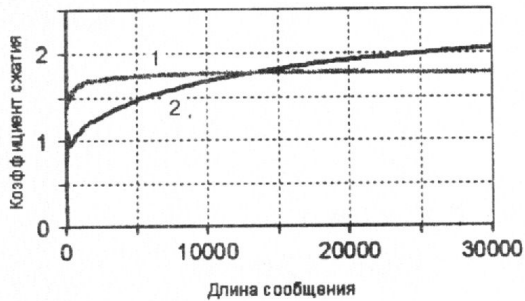


Рис. 2. Изменение коэффициента сжатия от длины прочитанного сообщения для алгоритмов Хаффмана (1) и LZW (2)

знака сжатие вообще не наблюдалось, а алгоритм Хаффмана уже показывал хороший результат. Это объясняется тем, что словарь LZW заполняется медленно, а алгоритм Хаффмана быстрее приспосабливается к статистике текста, поскольку он не учитывает корреляцию между знаками.

На рис. 2 представлен график изменения коэффициента сжатия от длины сообщения. При этом использовался текст на английском языке. Длина сообщения 30000 знаков.

График показывает, что по мере заполнения словаря алгоритма LZW увеличивается коэффициент сжатия. Видно, что со временем LZW опережает алгоритм Хаффмана по коэффициенту сжатия.

Рассмотрим изменение коэффициента сжатия при кодировании сообщения с нестационарными статистическими свойствами (рис. 3). Половина сообщения сформирована с одними статистическими свойствами, другая половина — с другими. Такая последовательность получена с помощью генератора псевдослучайных последовательностей. Для ее формирования были заданы матрицы условных вероятностей. Видно, что при изменении статистики сообщения коэффициент сжатия резко падает. Это позволяет находить изменение статистики путем отслеживания изменения коэффициента сжатия.

После анализа графиков был разработан адап-

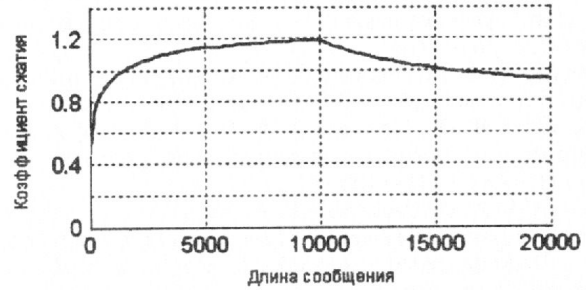


Рис. 3. Изменение коэффициента сжатия при кодировании сообщения с нестационарными статистическими свойствами

тивный алгоритм сжатия, принцип работы которого приводится ниже.

Вначале производится сжатие по адаптивному алгоритму Хаффмана и одновременно заполняется словарь LZW. С самого начала кодирования считаются частоты появления каждой фразы словаря LZW. При этом вычисляется коэффициент сжатия для каждого из алгоритмов. Кодирование по алгоритму Хаффмана продолжается до тех пор, пока алгоритм LZW не опередит его по коэффициенту сжатия. Затем в выходную последовательность записывается код переключения модели. Далее продолжается кодирование LZW до момента заполнения словаря. Когда словарь заполнен, формируются коды Хаффмана для всех элементов словаря. Далее продолжается кодирование LZW, но вместо индекса словаря в выходную последовательность записываются соответствующие коды Хаффмана. При дальнейшем кодировании отслеживается изменение коэффициента сжатия через каждые 100 знаков. Если коэффициент сжатия уменьшился, то следует произвести обновление модели и начать кодирование по адаптивному алгоритму Хаффмана.

Аналитическое проектирование алгоритмов сжатия сообщений должно заканчиваться оценкой их работоспособности и эффективности. В ряде случаев решить эти задачи аналитическими методами не представляется возможным.

Для исследования адаптивного алгоритма сжатия был проведен вычислительный эксперимент. С помощью генератора равномерных случайных величин были сформированы несколько псевдослучайных последовательностей с различными статистическими характеристиками: 1) последовательность независимых случайных величин с заданным распределением  $P_1(Z_i)$ , 2) последовательность коррелированных случайных величин с заданным двумерным распределением  $P_2(Z_{ij})$ , 3) последовательность коррелированных случайных величин с заданным трехмерным распределением  $P_3(Z_{ijk})$ , 4) последовательность коррелированных случайных

величин с неоднородной статистикой. Полученные последовательности затем сжимались. Значения коэффициентов сжатия приведены в таблице.

По результатам вычислительного эксперимента можно сделать вывод, что адаптивный алгоритм сжатия хорошо кодирует последовательности, у которых есть корреляция между знаками. Для некоррелированных последовательностей лучшим является алгоритм Хаффмана. По сравнению с алгоритмом LZW у адаптивного алгоритма коэффициент сжатия выше в среднем на 5 %. При больших объемах передачи информации это число является весомым.

Метод отслеживания изменения коэффициента сжатия можно применять в поточных алгоритмах сжатия как альтернативу фрагментированию. Хороший результат даст применение в схемах контекстуального моделирования PPM, поскольку это обеспечит хорошую адаптивность к резким изменениям данных.

1. Ватолин Д., Ратушняк А., Смирнов М., Юкин В. Методы сжатия данных. Устройство архиваторов, сжатие изображений и видео. — М.: Диалог-МИФИ, 2002.—384 с.
2. Дронь М. М., Малайчук В. П., Петренко О. М. Основи теорії захисту інформації: навч. посіб. — Д.: Вид-во Дніпропетр. уні-ту, 2001.—312 с.
3. Кадач А. В. Эффективные алгоритмы неискажающего сжатия текстовой информации: Дис. ... канд. физ.-мат. наук. — М.: Институт систем информатики имени А. П. Ершова, 1997.— (Машинопись).

---

ADAPTIVE ALGORITHM OF THE COMPRESSION  
OF NONSTATIONARY SIGN MESSAGES

E. V. Dzygin

The adaptive algorithm of the compression of nonstationary sign messages is developed. The change of the factor of the compression is experimentally investigated depending on length and type of a message. Some practical recommendations are derived concerning finding change statistics of the compressed message. A comparison of our algorithm of the compression with base ones is carried out.