

Комбинаторный способ сжатия «не сжимаемых данных» Филатов О. В.

Филатов Олег Владимирович / Filatov Oleg Vladimirovich - инженер-программист,
НТЦ Модуль, г. Москва

Аннотация: сжатие на один байт «не сжимаемых данных» с помощью поискового алгоритма, который путём перебора всех возможных значений и их сравнения с описанием несжатых данных восстанавливает утраченные при сжатии данные.

Abstract: compression by one byte «incompressible data» by using the search algorithm which by trying all possible values and comparing them with a description of uncompressed data restores data lost during compression.

Ключевые слова: сверхсжатие, сверхсжатые данные, уровень Колмогоровской сложности.

Keywords: supercompression, super compressed data, the level of Kolmogorov complexity.

Введение

Ныне действующие способы сжатия информации достигли своего предела. Этот предел формулируется в виде понятия «Не сжимаемого на один» байт данных. Доказаны теоремы, которые с математической строгостью объясняют существование предела сжатия от начального размера исходной информации. Но, тем не менее, в связи с развитием общего уровня знаний в соседних и родственных с математикой науках возникают новые научные концепции и идеи, применение которых в практической жизни приводят к трансформации самой математики путём ввода в математику новых понятий. Так, например, человечество создало компьютеры. Компьютерные алгоритмы и компьютерное моделирование стало не только изучаться в различных разделах математики, но и результаты работы компьютерных программ стали помогать в развитии математических идей и методов математических доказательств.

В этой статье описывается успешный компьютерный эксперимент, который позволил сжать «не сжимаемые на один» данные. Сжатие достигнуто не благодаря традиционному математическому аппарату, а благодаря отработанным компьютерным технологиям. Достигнутое сжатие вряд ли имеет практическое значение, оно слишком для этого мало. Но оно имеет концептуальное значение, так как показывает возможность преодоления предела сжатия данных за счёт идеи ассоциативного восстановления данных.

Примером ассоциативного восстановления данных служит восстановление старинных икон и картин. Когда утраченная часть изображения дорисовывается исходя из анализа оставшейся части иконы (картины) и общих культурных знаний историка – реставратора и искусствоведа.

Ассоциативное восстановление данных - это когда несуществующие данные создаются исходя из общего представления о целостном объекте. Недостающие данные восстанавливают по описанию когда-то существовавшего целостного объекта. То есть, вновь вводимые данные должны так вписаться в существующие данные, что имеющееся описание прежде существовавшего целого объекта будет хорошо описывать объединение оставшихся данных и вновь внесённых данных.

Присвоим не сжимаемому на единицу файлу (данным) единичный уровень Колмогоровской сложности. Демонстративное сжатие этого файла (данных) на один байт или же запись одного байта дополнительной информации в этот файл без увеличения его размера можно воспринимать как прохождение сквозь единичный уровень Колмогоровской сложности. Назовём сверхсжатием состояние данных, в которых преодолен единичный уровень Колмогоровской сложности, а сверхсжатые данные – это данные, хранящиеся в сверхсжатом файле. Впоследствии уровни сверхсжатия можно будет характеризовать числами, не равными единице.

В работе [1] был описан способ получения сверхсжатия, для использования которого требуется вычислительная мощь суперкомпьютера. В этой статье описан способ сверхсжатия, который можно проверить применительно к одному байту, на любом компьютере. И который не потребует знаний математического аппарата [1], так как в качестве его основной программы используется любой коммерческий компрессор, позволяющий сжимать и распаковывать без потерь файлы, и следящий за целостностью сжатой информации.

Возможность сжатия данных «в точку» в предлагаемом способе сверхсжатия данных исключается по двум причинам. Во-первых, из-за степенного роста требуемых вычислительных ресурсов и времени обработки данных, при увеличении степени сверхсжатия данных время обработки становится неприемлемо большим. Во-вторых, сверхсжатие данных в точку невозможно из-за невозможности выкинуть служебные данные компрессора. Так как целостность служебных данных компрессора обеспечивает возможность комбинаторного восстановления сверхсжатой информации. Исчезновение раздела со служебными данными компрессора (описание существовавшего ранее целостного объекта) приведёт к невозможности произвести комбинаторную разархивацию.

Основная часть

Процедура получения сверхсжатых данных. Сверхсжатие данных было получено для фотографии¹, рис. 1. Размер исходного файла фотографии 1 211 380 байт. Фотография дополнительно была сжата архиватором. Её размер после сжатия стал 1 180 772 байта. После сжатия архиватором файл стал «не сжимаемым на один» файлом (единичный уровень Колмогоровской сложности), который не мог быть далее сжат ранее известными способами.

Из этого файла был удалён один байт, сотый от начала файла. Номер удаляемого байта (сто) не был обоснован никакими соображениями, кроме его красоты. Содержимое выброшенного байта нигде не было сохранено. Размер сжатого файл стал равным 1 180 771 байт - на «ритуальный» один байт меньше, чем размер абсолютно не сжимаемого файла (1 180 772 байта). Поскольку этот файл (1 180 771 байта) можно разархивировать в первоначальном виде, то этот файл не является испорченным архивом (архивом с повреждёнными данными), а является файлом со сверхсжатыми данными.

Процедура восстановления данных из сверхсжатого файла. Ассистент перенёс сверхсжатые данные на свой компьютер. После чего на компьютере ассистента было создано 256 файлов (2⁸). В каждом из 256 файлов в сотую байтовую позицию был вставлен один байт. Теперь размер каждого из 256 файлов снова стал равным 1 180 772 байта. Содержимого одного файла от другого отличалось содержимым вставленного по адресу 100 байта. Каждый вставленный байт отличался от предшествующего вставленного байта на единицу. Значения вставленных 256 байтов: 0, 1, 2, ... 255.

Ассистент поочерёдно при помощи разархиватора пытался разархивировать каждый из 256 файлов. В 255 случаях процесс разархивации не запускался, и выдавалось сообщение об ошибке архивных данных. Никаких разархивированных файлов не возникло.

В одном случае (из 256) процесс разархивации был произведён штатно, без сообщения об ошибке и с образованием разархивированного файла. Для данной конкретной фотографии это произошло при значении вставленного байта, равному 66. Файлы, содержащие байты с числами не равными 66, не смогли разархивироваться.

Таким образом, из файла со сверхсжатыми данными (1 180 771 байта) был получен файл с абсолютно не сжимаемыми данными (1 180 772 байта), из которого было восстановлено исходное изображение, рис. 1.



Рис. 1

Обсуждение

«Поисковый алгоритм» проявляет себя в рассмотренном выше восстановлении данных картинки на этапе разархивации. Поиск заключается в обнаружении такого числа в вставляемом в файл байте, при котором обсчёт параметров данных в файле совпадёт с прежнем обсчётом этих параметров. Прежний обсчёт хранится как часть «не сжимаемого на один» файла.

Можно выделить два типа «не сжимаемых на один» файлов: файлы с сжатой информацией и файлы, образованные из случайных бинарных событий (не несущих ни какой информации). При попытке сжатия достаточно длинных случайных бинарных файлов (данных), архиватор увеличивает размер «сжатого» файла, добавляя к ним свою информацию, и вместо уменьшения размера файла получается, наоборот, рост его длины. Поэтому для получения сверхсжатого файла случайных бинарных данных необходимо выбросить при сверхсжатии участок файла, длина которого превысит длину добавленной архиватором информации. Очевидно, что декомпрессия сверхсжатого файла случайных бинарных данных методом комбинаторного перебора потребует значительных аппаратных и временных ресурсов.

Вывод

В статье был продемонстрирован способ получения сверхсжатых данных и процедура комбинаторной разархивации (восстановления) сверхсжатых данных до состояния «несжимаемых далее данных».

Сверхсжатие достигается путём прямого уничтожения части информации. Восстановление уничтоженной информации производится путём включений вновь придумываемой информации в

¹ Дизайнер арт-объектов фотографии, фотограф-ассистент Филатов Илья Олегович

имеющуюся информацию (объединение информации) и сравнения объединённой информации с описанием, каким была информация до потери своей части.

Восстановление осмысленных данных (например, изображения) отличается от восстановления бессмысленных данных полученных при генерации случайных бинарных событий тем, что применение компрессора к изображению уменьшает размер итогового файла с картинкой, а применение компрессора к достаточно большому файлу со случайными бинарными данными создаёт итоговый файл ещё большего размера.

Литература

1. *Филатов О. В.* статья «Описание распределения составных событий и их мизесовских частот через число возможных исходов. Механизм сжатия некоторых «не сжимаемых на один» последовательностей. «Проблемы современной науки и образования», № 9 (39), 2015 г.