



ОБЪЕДИНЕННЫЙ
ИНСТИТУТ
ЯДЕРНЫХ
ИССЛЕДОВАНИЙ
ДУБНА

3413/82

9/8-82

P10-82-304

А.А.Бялко, Н.Г.Волков, В.М.Цупко-Ситников

МЕТОДЫ СЖАТИЯ ДАННЫХ
В ЗАДАЧАХ ЭКСПЕРИМЕНТАЛЬНОЙ
ЯДЕРНОЙ ФИЗИКИ

Направлено в сборник "Прикладная ядерная
спектроскопия"

1982

1. ВВЕДЕНИЕ

Проблемы, вызванные увеличением объемов информации, можно разделить на три основные группы. Это проблемы кодирования /передачи/, хранения и обработки данных. В современной физике присутствуют все три компонента работы с данными, но самые серьезные трудности связаны с двумя последними - хранением и обработкой большого объема информации. Вопросам сжатия данных и оптимизации процессов работы с информацией посвящено большое количество работ. Они касаются различных аспектов этой проблемы и предлагают множество разнообразных подходов. В этом плане нам представляется полезным рассмотреть методы уменьшения объемов данных в области физики низких энергий и указать на способы оценки их эффективности исходя из положений теории информации.

В настоящее время в теории информации существует несколько различных подходов оценки количества информации, которые имеют разные уровни разработки. Не касаясь особенностей этих подходов, связанных чаще всего с областью их применения, мы предлагаем использовать те информационные методы, которые были бы естественными для физика-экспериментатора /отвечали бы его интуитивным и практическим представлениям/.

Наибольшее развитие, особенно в теоретическом плане, получило решение задачи кодирования данных, что было вызвано потребностями развития средств связи. Именно при решении этих задач К.Э.Шенноном были заложены основы теории информации /1/ /классическая теория информации/. Следует отметить, что выводы классической теории информации носят статистический характер. Поэтому, применяя ее для практических задач, необходимо помнить, что все расчеты справедливы лишь для потока сообщений, а для конкретного сообщения могут быть неверны.

Среднее количество информации, содержащееся в дискретном сообщении, равно согласно Шеннону

$$I_{\text{ср}} = -n \sum_{k=1}^m p_k \log p_k, \quad /1/$$

где n - общее число символов сообщения; p_k - вероятность появления k -го символа; m - число символов, из которых состоит алфавит сообщения.

Максимум I_{cp} достигается /с учетом того, что $\sum_{k=1}^m p_k = 1$ /
при $p_k = \frac{1}{m}$ для всех K . Количество информации при этом равно

$$I_{cp} = n \log m. \quad /2/$$

Эта формула имеет самостоятельное значение /она была получена Р.Хартли в 1928 году/. Выражение /2/ указывает теоретический предел, который может быть достигнут за счет применения экономного кодирования. Оно определяет и способ увеличения информативности: переход к кодированию, при котором все символы появлялись бы равновероятно. В классической теории информации мера Хартли является той величиной, которая должна сохраняться при преобразованиях.

Однако эта мера количества информации не отвечает нашим обычным представлениям об информативности сообщения. В таком сообщении как, например, экспериментальный спектр гамма-излучения определенного ядра /точнее - распределение амплитуд импульсов от детектора гамма-излучения/, насыщенность его информацией, представляющей для нас интерес, может быть очень далека от предельно возможной в шенноновском определении.

Обработка экспериментальных распределений сигналов от детектора гамма-излучения представляет собой процесс преобразования информации с целью получения полезных, интересующих нас данных - истинного гамма-спектра, то есть энергий и интенсивностей входящих в него линий. В результате обработки информации образуется новый набор данных, как правило, меньшего объема, чем исходный. В общем случае обработка является сжатием данных, причем величина сокращения их может значительно превышать пределы, которые дает информационный подход Хартли-Шеннона.

Очевидно, что в отношении первичных экспериментальных данных необходим иной подход к оценке информации, учитывающей не только вероятностные характеристики появления отдельных символов в сообщении, но и структурные свойства данных, которые представляют основную ценность, например, в спектрометрии.

Таким подходом является, по мнению авторов, алгоритмический подход А.Н.Колмогорова^{/2/}. Согласно А.Н.Колмогорову наиболее содержательным является понятие о количестве информации в одном объекте (X) о другом объекте (Y). Свою количественную оценку информации А.Н.Колмогоров строит на понятии сложности объекта, которая определяется как минимальная длина программы, необходимая для воспроизведения объекта.

Для определения информации в объекте X об объекте Y им вводится понятие относительной сложности, которая является минимальной длиной программы получения Y из X. Непосредственно информация в объекте X об объекте Y определяется как разность сложности объекта Y и его относительной сложности к объекту X.

Очевидно, что такое определение понятия информации позволяет подойти к оценке предельных возможностей и эффективности различных методов сжатия исходя из более близких, практически используемых нами представлений об информативности получаемых экспериментальных данных. К сожалению, в настоящее время уровень программирования не позволяет создать универсальные количественные оценки информации для случая обработки. Поэтому при использовании определенных методов обработки информация оценивается либо с помощью элементов обычного алгоритмического характера, либо другими информационными подходами^{/49/}.

Возможность получения оценки величины информации в каком-либо объекте позволяет лишь указать предел сокращения объема данных. Реализовать это сокращение можно, применяя различные методы сжатия данных /сжатия информации, сокращения избыточности, встречаются и другие термины/.

В любом случае сжатие приводит к потере информации^{/2/}. Направленное на определенные характеристики данных, сжатие сводится к исключению ненужной в данном случае информации и выделению полезной в каком-то аспекте ее части. Без этого сжатие экспериментальных данных невозможно. Лучший /предельный/ вариант сжатия - полная обработка данных, но она не всегда возможна в ходе эксперимента, и сам характер обработки может зависеть от полученного результата. Обычно применяются быстрые алгоритмы сжатия с целью уменьшения объема данных при их хранении до окончательной обработки. При этом предпочтительно сжатие, являющееся частью обработки, ее начальным этапом, который может быть впоследствии продолжен без каких-либо дополнительных действий.

Эффективность методов сжатия может рассматриваться не прямо по отношению, например, к оптимальному алгоритму обработки, а в зависимости от конкретных условий эксперимента, возможностей аппаратуры, используемой для регистрации и анализа данных. Например, при конкретных обстоятельствах сжатие объема данных в несколько раз может быть обязательным условием выполнения эксперимента, хотя и значительно усложняющим последующую обработку результатов.

Имеется ряд работ, систематизирующих существующие методы сжатия данных^{/3-9/}. В работе Ю.Б.Ольховского и др.^{/3/} основой классификации методов сжатия данных служит возможность их восстановления /точного или приближенного/ после сжатия. Возможность точного восстановления определяется как обратимость метода, приближенность - как квазиобратимость. Отсутствие таких возможностей можно назвать необратимостью.

В^{/4/} не оговаривается специально наличие систематизации, однако описание методов подразделено следующим образом: методы

повышения эффективности использования ячеек ЗУ, методы отбора полезной информации, ортогональные преобразования и ассоциативные методы. В работах^{/5-9/} акцент делается на технических сторонах реализации алгоритмов. Библиография работ по сжатию данных /опубликованные до 1971 г./ имеется в специальной подборке^{/10/}.

Использование методов сжатия данных предполагает, что количество информации, первоначально содержащееся в данных, не будет сильно изменено /по какому-то выбранному критерию/. Поэтому в настоящей работе алгоритмы сжатия данных подразделяются на три группы в зависимости от того, какие теоретические положения необходимо применить для оценки содержащейся в данных информации: подход Шеннона-Хартли /1/, положения А.Н.Колмогорова, относящиеся к алгоритмам обработки /2/, либо к алгоритмам хранения информации /3/. В первую группу входят алгоритмы, основанные на кодировании и обладающие лишь средними по массивам показателями. Во вторую входят алгоритмы с элементами обработки. Третья группа объединяет алгоритмы для хранения информации в преобразованном виде. Для этой группы будет введена оценка количества информации, основанная на понятии "сложности" А.Н.Колмогорова.

2. МЕТОДЫ СЖАТИЯ ДАННЫХ, ОСНОВАННЫЕ НА СТАТИСТИЧЕСКИХ ЗАКОНОМЕРНОСТЯХ

Применение методов сжатия данных, основанных на статистических закономерностях, широко распространено в вычислительной технике, связи, экспериментальной физике и других областях. Это обусловлено простотой вычислений для этой группы методов, многие из которых выполняются всего несколькими арифметическими операциями. Вычисления проводятся обычно с небольшим объемом данных, за счет чего и возможно большое быстроедействие. При этом, однако, упускаются те свойства, которые могут дать наибольший эффект сокращения данных.

Следует помнить также о статистическом характере этих методов. Любой из методов первой группы может не давать положительного эффекта или даже давать отрицательный для конкретного случая. Использование таких методов требует, чтобы экспериментатор представлял себе /хотя бы очень приблизительно/ характер данных, которые будут получены.

По отношению к числовым данным теория информации по Шеннону рекомендует находить такую форму представления, которая имела бы одинаковую /среднюю/ длину единиц /как правило, данные выражаются в двоичных единицах - битах/.

Существует несколько реализаций этой задачи. Наиболее простое и часто используемое - представление чисел в форме с плавающей запятой. Количество бит, отводимое под мантиссу и порядок, определяется диапазоном представления чисел и точностью вычислений, причем общее количество бит всегда меньше, чем у максимального /минимального/ числа в представлении с фиксированной запятой. Операции с числами в таком виде связаны с ошибкой. Сами числа получают одинаковые относительные ошибки, преобразуясь к такой форме. Характер погрешностей, возникающих при работе с числами в форме с плавающей запятой, проанализирован в книге Маккракена и Дорна^{/11/}. Преобразование к этой форме целых чисел, получаемых в результате физических измерений, рассматривается в^{/12/}, где дан анализ и других простых способов оптимизации хранения чисел. Выравнивание требуемой памяти для каждого числа может быть проведено следующим образом. Все числа /целые/ делятся на константу, большую единицы, результат округляется до определенного числа знаков. Такой способ /масштабирование чисел/ приводит к меньшим относительным ошибкам для больших чисел. Следовательно, такое представление целесообразно в тех случаях, когда интерес представляют большие числа. Масштабирование может производиться на заранее выбранную константу или динамически^{/12/}. Примером динамического масштабирования может служить накопление интенсивностей вместо абсолютных значений^{/13/} в спектрометрии.

Ко многим задачам физики применимы методы экономного кодирования, когда существует необходимость передачи чисел /кодов/. Эти методы хорошо изучены в теории связи и решение задачи оптимального кодирования сводится к техническим моментам^{/14/}. Простейшим кодом, который позволяет сократить объем сообщения, является код Шеннона-Фано. Он ставит в соответствие чаще встречающимся символам меньшее число двоичных знаков. Более сложный вариант подобной кодировки - код Хофмана^{/15/}.

При работе с длинными цепочками символов разработаны специальные методы кодирования, основанные на выделении повторяющихся частей^{/15/}. Все эти методы кодирования по своей конструкции позволяют точно восстанавливать первоначальную информацию.

Получение распределений в виде гистограмм связано с тем, что ячейка анализатора должна иметь размер максимально возможного значения. Выравнивание значений отсчетов в ячейках анализатора /что согласуется с теорией Шеннона/ может быть достигнуто методом стохастического счета^{/4,16/}. Этот метод позволяет накапливать вместо распределения функцию квадратного корня из него. Основа метода - в том, что запись события в канал анализатора производится не регулярно, а с некоторой вероятностью, которая выбирается^{/4/} для удобства вычислений сначала равной единице, пока значение отсчета в канале не достигает опреде-

ленной величины, а затем обратно пропорциональной накопленным отсчетам. Реализуется подобная процедура с помощью псевдослучайных чисел. Если псевдослучайное число /от 0 до 1/ меньше вычисленной вероятности записи^{/4/}, то запись производится, в противном случае нет. Естественно, данный метод не позволяет точно восстанавливать распределение, однако, как показано в^{/16/}, он не вносит дополнительных искажений и даже устраняет часть имеющихся. Приведенный алгоритм позволяет использовать анализатор с ячейками вдвое меньшей длины.

Следующая подгруппа алгоритмов сжатия данных основана на устранении корреляций между соседними значениями. Коррелированность, так же как и неравномерность статистического распределения, по Шеннону, снижает количество информации, увеличивает избыточность данных. Наиболее полное и современное описание подобных методов, относящихся в первую очередь к передаче непрерывных сообщений, дается в работе В.А.Свириденко^{/5/}.

Часть алгоритмов, устраняющих корреляции между соседними значениями сообщения, основана на вычислении разности соседних значений. Это так называемые разностные /дельта-/ представления. Существует несколько вариантов дельта-алгоритмов с использованием предсказания. Более подробно они описаны в работе^{/5/}, там же содержатся ссылки на оригинальные работы. Отметим только, что применительно к двумерной информации /телекино изображения/^{/5,17/} оптимальное кодирование - межкадровое кодирование. В последнем случае корреляции устраняются между последовательными кадрами, которые в большинстве очень близки^{/17/}.

Цифровой метод, являющийся частным случаем разностных алгоритмов, приводится в работе^{/18/}. Автор предлагает при анализе распределений вместо обычного спектра накапливать его численную производную. Для этого при поступлении события с адресом K в K -ю ячейку заносится $+1$, а в $K-1$ -ю заносится -1 . Необходимо зарезервировать одну ячейку для накопления общего числа поступивших событий, что позволит восстановить спектр.

Из-за конечного разрешения спектров разность соседних значений спектра всегда бывает меньше самого значения, что позволяет достигать сокращения данных. На сравнительно гладком гамма-спектре было получено сокращение вдвое требуемой памяти.

Развивая метод, можно воспользоваться следующими численными производными. Для регистрации второй численной производной необходимо при поступлении события K записать $-1, 2, -1$ в ячейки $K+1, K, K-1$ соответственно и зарезервировать две ячейки для восстановления спектра. Но, как показало исследование второй производной^{/18/}, в этом случае начинает возрастать доля случайной компоненты. Поэтому абсолютные значения высоких производных начинают расти.

Эффективность методов сжатия данных, описанных в этом разделе применительно к задачам ядерной физики, невысока. Как правило, величина сокращения данных чаще всего равняется двум. Это соответствует и теоретическому пределу, который дает классическая теория информации для такого типа данных. Многие из перечисленных методов могут применяться не как основные, а как вспомогательные для более сложных подходов.

Во всех приведенных методах использовались по отдельности устранение неравномерности или коррелирование данных. Существует путь увеличения эффективности вероятностных /статистических/ методов за счет одновременного устранения обоих видов избыточности.

3. МЕТОДЫ СЖАТИЯ ДАННЫХ, ИСПОЛЬЗУЮЩИЕ ЭЛЕМЕНТЫ ОБРАБОТКИ ИНФОРМАЦИИ /ОТБОР ИНФОРМАЦИИ ПО ОПРЕДЕЛЕННЫМ ПРИЗНАКАМ/

Большинство методов этой группы используют начальные, чаще всего наиболее простые и быстрые, этапы обработки информации с целью выделения наиболее существенной ее части и сокращения объема памяти для хранения данных. Эти методы обычно применяют, когда известны алгоритм полной обработки и характер информации, которая должна быть получена в результате ее применения.

Очень часто, если не используется сжатие, эти методы легко осуществляются экспериментатором визуально на обширных одномерных или многомерных полях экспериментальных данных. Но такой подход не всегда возможен из-за необходимости очень больших объемов памяти и достаточно совершенных средств представления информации.

Методы этого раздела основаны на учете определенных свойств данных и поэтому отличаются большим разнообразием. Структура и эффективность метода меняется в зависимости от заинтересованности пользователя тем или иным аспектом данных. Среди этих методов можно выделить два направления: получение некоторых функций от большого числа значений регистрируемых переменных и выделение областей с "полезной" информацией.

Простейшим примером первого типа алгоритмов может служить замена дублирования /обычный прием для обеспечения надежности/ проверкой четности при считывании информации. Дополнительный бит служит для записи показателя четности числа единиц /в двоичном представлении/ и достаточно надежно контролирует качество считывания и намного экономичней дублирования. Существуют и более сложные методы подобного контроля.

Более сложное преобразование данных первого направления - метод контурного кодирования, которое особенно удобно в дву-

мерном случае^{/21/}. Суть метода состоит в представлении данных изолиниями, как например, в географических картах. Хранение координат изолиний вместо всего объема данных может быть во многих случаях экономично, тем не менее задача получения такого представления достаточно сложна^{/21,22/}.

Это еще осложняется и тем, что в реальном случае приходится иметь дело с искаженными данными. Поэтому в физике низких энергий, особенно в процессе накопления, использование подобных алгоритмов пока невозможно из-за их ограниченного быстродействия.

Развитие такого подхода может быть целесообразно в физике высоких энергий. При анализе двумерных массивов /фотографий, телеизображений/ большая часть чисел составляет нули. Отличные от нуля значения располагаются вдоль треков частиц, определение параметров которых является задачей обработки. Специфический вариант такой обработки - алгоритм слежения по треку^{/23/}. Применение этого алгоритма в процессе приема информации позволяет сразу записывать координаты треков и тем самым сокращать объем данных в десятки раз^{/23/}.

Для произвольных данных метод контурного кодирования во многом применим только при наличии "быстрых" алгоритмов. Работы в этой области существуют^{/24/}.

Работу многих алгоритмов принято оценивать субъективно. Это, как правило, те алгоритмы, для которых получателем информации служит человек. В таких случаях можно воспользоваться свойствами человеческих органов слуха и зрения и значительно упростить данные^{/5,7,9,17,21,25/}. В физике эффективность сокращения данных может быть увеличена за счет применения такого типа алгоритмов, если информация предназначена для визуального контроля, а не для прецизионной обработки.

Другая большая группа методов сжатия данных основана на выявлении участков существенной /по каким-то критериям/ информации и отбрасывании всей остальной. По такому принципу работают апертурные алгоритмы, которые применяются при передаче информации. Их подробное описание, а также полную библиографию по этому вопросу, можно найти в книге В.А.Свириденко^{/5/}.

Апертурные алгоритмы выделяют "информативные выборки" из временных последовательностей данных при помощи аппроксимации и сравнения результатов аппроксимации с реальными значениями. Если аппроксимация слабо /по некоторому критерию/ отличается от текущего значения, то запись или передача этого значения не нужна. Эти алгоритмы сильно различаются по виду аппроксимирующей функции, способу аппроксимации и другим деталям^{/5/}.

Применение апертурных алгоритмов включает в себе противоречие: алгоритм "работает" только при возможности хорошей аппроксимации, но если она известна, то информация лишена смысла.

Это особенно хорошо видно при аппроксимации константой /так называемый предсказатель нулевого порядка /ПНП//. Выход из этого противоречия - в применении подобных алгоритмов к данным, состоящим из больших участков хорошо аппроксимирующихся областей и малых участков другого рода. В работе^{/6/} приводится характерный пример удачного использования ПНП. Алгоритм использовался для передачи данных о радиоизлучении Солнца, которое характеризуется вспышками и длительными спокойными периодами. При этом объем данных удалось сократить в десятки раз с применением ПНП по сравнению с непрерывной передачей^{/6/}.

Интересный пример использования линейного предсказателя для нужд физики высоких энергий описан в работе^{/47/}. В этом случае линейный предсказатель применялся для нахождения координат треков частиц на изображениях, полученных со стримерных или пузырьковых камер.

Так же эффективны и методы окон при измерении статистических распределений, характеризующихся наличием областей пиков. Можно найти аналогию между выделением пиков в распределениях и участками активности в приведенном примере. Существенным отличием является только процесс выборки "полезных" или активных участков. При передаче информации не существует другого способа отбора, кроме основанного на анализе предыдущей части сообщения. Для отбора участков распределений в методе окон их может быть несколько. Можно использовать априорный отбор с участием экспериментатора или автоматический отбор^{/26,28/}. С технической точки зрения метод окон реализуется на уровне аналоговой электроники и на уровне цифровой информации^{/26,27/}.

Быстродействие метода окон определяется скоростью сортировки событий по окнам, а эффективность - соотношением "полезных" и "бесполезных" участков. Для конкретных случаев в области гамма-спектрометрии^{/26,27,28/} сокращение данных методом окон находилось на уровне десятков или даже сотен. Подробный разбор методов выделения полезной информации содержится в^{/4/} и^{/28/}, там же можно найти ссылки на оригинальные работы.

Несмотря на высокую эффективность, метод окон используется лишь на определенном уровне измерительной техники. Его применение связано с некоторыми трудностями. В варианте с априорной селекцией возникает необходимость подготовки большого объема данных до эксперимента, повышаются требования к аппаратуре, усложняются изменения в ходе эксперимента. Кроме того, часто положение всей полезной части информации неизвестно.

Если положение участков задается экспериментатором по ходу эксперимента, то даже в самом оптимальном случае - разметка на дисплее световым карандашом - это требует больших затрат квалифицированного ручного труда.

Автоматический вариант отбора требует обработки в процессе накопления информации. Эта задача в физике низких энергий достаточно сложна и должна решаться быстро в условиях неполной информации.

По мере развития возможностей регистрирующей аппаратуры метод окон все реже используется непосредственно в процессе накопления информации. С другой стороны, он стал широко применяться в различных программах обработки многомерных данных^{/29,30/}.

Недостатки метода окон стимулировали поиск методов, которые позволили бы просто и без вмешательства человека отбирать информацию. Одним из таких подходов является ассоциативный метод. Его можно рассматривать как метод окон с единичными размерами последних, которые называются дескрипторами. Существенное отличие метода окон от ассоциативного заключается в следующем: в первом координаты окон занимают незначительную часть памяти, во втором - половину; в первом поиск окон - простая задача, а для второго - наиболее трудоемкая. Развитие ассоциативного метода можно проследить по книге В. Соучека^{/28/}, в которой приведено изложение более ранних работ автора. Обзор, включающий и отечественные работы, дан в^{/4/}.

Самый простой вариант работы ассоциативного алгоритма состоит в следующем: поступающие коды событий анализируются на наличие в списке дескрипторов. Если дескриптор находится в списке, то в соответствующую данному дескриптору ячейку добавляется единица. Если нет - вводится в список новый дескриптор, а в соответствующую ему ячейку записывается единица. Этот вариант обладает двумя недостатками: большое число сравнений и вероятность выбора неправильного дескриптора при наличии фона.

Первый недостаток может быть устранен программно^{/28/} или при помощи специализированных устройств^{/31/}, в которых проводятся одновременно сравнения приходящего адреса с группой адресов из списка дескрипторов. Принципиальная разрешимость задачи в отсутствие фона означает, что ассоциативный анализ является самым простым и точным способом регистрации распределений в границах сложной формы при отсутствии значений вне границ.

При наличии фона, который может на отдельных участках превышать эффект /спектры частиц и гамма-квантов/, ассоциативный алгоритм должен включать сложный статистический анализ, что приводит к потерям и быстродействия, и сокращения памяти. По этой причине ассоциативный метод не нашел широкого применения в области физики низких энергий, хотя на отдельных типах данных он давал сокращение объема памяти более чем в десять раз^{/28,32/}.

Перечисленные в данном разделе методы сжатия данных обладают наиболее высокой эффективностью, которая достигается за счет направленности их на определенный аспект информации. Сле-

дует отметить, что большая часть методов этого раздела необратима, то есть не позволяет восстанавливать первичный вид информации, исключает повторные обработки другими методами. Это предъявляет особые требования к выбору и использованию таких методов.

4. СЖАТИЕ ДАННЫХ С ПОМОЩЬЮ ПРЕОБРАЗОВАНИЙ

Методы преобразования данных можно сгруппировать так: методы специальных преобразований, преобразования координат, разложение по системам функций.

Специальные преобразования данных с целью сокращения их объема требуют хорошего знания конкретной задачи.

Здесь можно отнести различные преобразования координатной плоскости поля данных в координатную ось с помощью сложения, умножения и т.д. первичных кодов при двумерных измерениях^{/19/}. При этом следует подбирать преобразования, которые способствуют проявлению таких структур данных, которые представляют наибольший интерес в данном измерении и позволяют сократить общий объем информации.

Сокращая объем данных в десятки или даже сотни раз, вместе с тем можно сохранить возможность качественной оценки всего двумерного распределения по его одномерному образу^{/19/}. Применение данной методики оставляет возможность и количественного анализа двумерного спектра с использованием метода решения некорректных задач^{/20/}. При этом необходимо использовать несколько одномерных интегральных образов одновременно.

Универсальных рецептов, определяющих, какие преобразования следует использовать, пока нет, и лучше всего проиллюстрировать принцип на примере. При определении характеристик осколков делящихся веществ /Шмит и др.^{/33/} / одновременно и независимо определялась энергия осколков деления E и время пролета T . Накопление двумерного спектра, который имел 10^5 каналов, требовало огромного количества событий деления. Вместе с тем величины E и T не являются независимыми. Величина ET^2 пропорциональна массе осколка. Таким образом, двумерный спектр обуславливается одномерным массовым распределением, а его регистрация позволяет на два порядка уменьшить число каналов, улучшить статистику.

Очень близки таким преобразованиям преобразования координат. Зная общую тенденцию поведения измеряемых величин, можно подобрать такие координаты, в которых наиболее полно проявляется структура данных. Известно, например, что сложные суммы экспонент лучше анализировать при разбиении отрезка на интервалы экспоненциально изменяющейся длины^{/34/}. В нейтроно-

графии существуют методы^{/35/}, сокращающие объем данных путем преобразования пространства измерений.

Поскольку подобные преобразования предполагают глубокое понимание процессов, лежащих в основе получаемых данных, развития таких методов следует ожидать от экспериментаторов.

Наиболее универсальные методы сжатия данных основаны на разложении по системам линейно независимых функций. Для простоты вычислений система функций, как правило, выбирается ортогональной.

Эта группа методов сжатия данных объединяет методы, основная цель которых - сохранение информации близкой /по установленной норме/ к исходной. Норма строится на всем множестве исходных и восстановленных после преобразования величин. Такой подход подразумевает полное сохранение структуры данных и возможность их последующей многократной обработки различными методами. В отличие от сжатия с использованием элементов обработки, это более универсальный и объективный подход.

Применение ортогональных преобразований разнообразно: это применение по типу обрезания фильтра^{/36/}, блочное квантование^{/5/} /в этом случае выигрыш достигается благодаря выбору уровня дискретизации преобразованных данных/, пороговый отбор - сохранение компонент, превышающих определенную величину^{/5,37/}.

Применяемые ортогональные базисы различны и обусловлены сложностью реализации. Это преобразование Фурье^{/38/}, Уолша-Адамара^{/37/} и близкие им^{/39/}, Хаара^{/40/}, пилообразные функции^{/41,48/}, полиномы Эрмита^{/42/}, преобразование Корунена-Лоева-Пугачева^{/36,43/} и др. Опыт применения ортогональных преобразований подсказывает введение в качестве оценки количества информации после преобразования суммарной длины программы обратного преобразования и величин, необходимых для такого преобразования. Подобная оценка согласуется с алгоритмическим подходом А.Н.Колмогорова и близка по величине к введенному им понятию "сложность"^{/2/}.

При использовании всех ортогональных преобразований, не имеющих "быстрых" алгоритмов реализации, необходимо хранить либо программу получения ортогональных функций, либо значения функций. По быстройдействию преобразования двумерного массива $N \times N$ требуют $K_1 N^4$ операций в общем случае и $K_2 N^2 \log N^2$ - при наличии "быстрого" алгоритма, K_1 и K_2 - коэффициенты пропорциональности, не зависящие от размеров массива. Если при поступлении события накапливаются по одному, то, воспользовавшись свойствами дельта-функции на сетке, задачу можно значительно упростить. Но восстановление данных остается связанным с огромным объемом вычислений.

Чаще всего, если используются преобразования Фурье, Уолша и подобные им, они выявляют самые общие черты структуры данных. Исключение составляет, например, кодирование звуковых сиг-

налов, когда базис совпадает с аппаратурной линией^{/44/}. Существуют и преобразования, которые отражают и более конкретные свойства. Так, в двумерном случае многими оптимальными свойствами обладает разложение по сингулярным значениям^{/40,45/}.

Свойства разложения по сингулярным значениям обусловлены его определением: любая действительная матрица A представима в виде

$$A = U \sigma \cdot V^T, \quad /3/$$

где U и V - матрицы, столбцы которых являются собственными векторами матрицы $A^T A$ и AA^T соответственно; σ - собственные значения этих матриц /или сингулярные значения/. Эта формула может быть записана иначе:

$$A_{ij} = \sum_{k=1}^R \sigma_k U_{ik} V_{jk}, \quad /4/$$

где R - ранг матрицы. Применяя подобные преобразования к двумерным распределениям, в частности к спектрам гамма-гамма-совпадений^{/46/}, оказалось возможным благодаря уменьшению R не только сохранить спектры близкими по норме к исходным, но и достичь эффекта фильтрации помех. При этом достигается 8-кратное сокращение объемов данных. Это позволяет хранить спектры совпадений в виде U, V, σ непосредственно в оперативной памяти больших ЭВМ. Восстановление одного элемента спектра требует $2 \times R$ операций умножений по выражению /4/, что значительно быстрее сортировки с внешних носителей и не уступает показателям "быстрых" алгоритмов. С точки зрения сжатия данных эти свойства особенно ценны для данных с наличием симметрии по оси матрицы. Кроме того, это преобразование оптимально по критерию объема данных, программы восстановления и быстройдействию, которое для восстановления не хуже показателя "быстрых" алгоритмов.

Алгоритмы сжатия данных, основанные на ортогональных преобразованиях, имеют еще два преимущества. Во-первых, они допускают применение в качестве дополнительных - алгоритмов первой группы. Во-вторых, при правильном использовании они "улучшают" данные, избавляя их от части помех /как пример можно привести преобразование Фурье, при использовании которого уменьшается высокочастотный шум/.

Правильный подбор алгоритма и ортогонального базиса дает возможность, как правило, добиться этими методами сокращения объемов памяти до 10 раз. При этом важная роль принадлежит специфике исследования и получения данных. Кроме того, следует обратить внимание на одно обстоятельство: целесообразно искать прямые пути получения из сжатых данных интересующей нас инфор-

мации. Это не всегда очевидный путь, и часто сжатые данные восстанавливают в первоначальную форму и затем проводят обработку обычным способом, что значительно усложняет путь получения окончательной информации. Свернутые экспериментальные данные при правильном выборе преобразования могут быть достаточно близки к конечному результату и это надо использовать.

5. ЗАКЛЮЧЕНИЕ

В настоящей работе авторы привели оценку методов и приемов сокращения объемов данных с позиций физика-экспериментатора. Мы намеренно мало уделяли внимания подробному описанию алгоритмов и акцентировали внимание на тех моментах, с которыми приходится сталкиваться при выборе и использовании методов сжатия.

Важным мы считаем и соединение практических методов с основами теории информации в различных ее вариантах.

Наиболее перспективна, как нам кажется, в настоящее время группа методов, связанная с преобразованием данных. Эти методы имеют высокие показатели и в эффективности сжатия, и по быстродействию, позволяют сохранять информацию, близкую к исходной. Особенно выделяются в этой группе алгоритмы с базисом преобразования, определяемым самими данными.

Метод разложения по сингулярным значениям может быть реализован с тем же количеством операций, что и у "быстрых" алгоритмов ортогональных преобразований⁴⁸. В дополнение отметим, что ортогональные преобразования могут использовать методы первой группы как вспомогательные и сами могут быть в роли вспомогательных, например, для метода окон и ему подобных.

В заключение авторы выражают благодарность Г.А.Ососкову за полезные обсуждения работы при подготовке ее к печати.

ЛИТЕРАТУРА

1. Шеннон К. Работы по теории информации и кибернетике. ИЛ, М., 1963, с.830.
2. Колмогоров А.Н. Проблемы передачи информации, 1965, т.1, вып.1, с.3-11.
3. Ольховский Ю.Б., Новоселов О.Н., Мановцев А.П. Сжатие данных при телеизмерениях. "Сов. радио", М., 1971, с.304.
4. Маталина Л.А. Электронные методы ядерной физики. Атомиздат, М., 1973, с.520.
5. Свириденко В.А. Анализ систем со сжатием данных. "Связь", М., 1977, с.183.

6. Ключев В.И., Свириденко В.А. Передача сообщений с повышенной информативностью /Радиоэлектроника и связь/, "Знание", М., 1973, №8.
7. Кортман С.М. ТИИЭР, 1967, т.55, №3, с.88.
8. Хочман Д. ТИИЭР, 1967, т.55, №3, с.21.
9. Андрус С.А. ТИИЭР, 1967, т.55, №3, с.25.
10. Wilkins L.C., Wintz P.A. Bibliography on Data Compression, Picture Properties and Picture Coding. IEEE, 1971, tr. IT-17, No.2, p.180-197.
11. Мак-Кракен Д., Дорн У. Численные методы и программирование на фортране. "Мир", М., 1977, с.584.
12. Culhane J.L., Nettleship R. IEEE, 1968, tr.NS-15, No.2, p.3.
13. Hooton I.N., Best G.C. NIM, 1967, vol.56, pp.277-284.
14. Штарьков Ю.М. В кн.: Кодирование в сложных системах. "Наука", М., 1974.
15. Dishon Y. Computer Design, 1977, No.4, pp.85-90.
16. Цифровое телевидение /под ред. М.И.Кривошеева/. "Связь", М., 1980, с.263.
17. Inouye T. NIM, 1971, vol.91, p.581.
18. Waugh J.V.S. NIM, 1968, vol.58, pp.293-297.
19. Громов К.Я. и др. ПТЭ, 1974, №4, с.64.
20. Тихонов А.Н., Арсенин В.Я. Методы решения некорректных задач. "Наука", М., 1974.
21. Грехель Д.Н. ТИИЭР, 1967, т.55, №3, с.102.
22. Бьюси Р. ТИИЭР, 1970, т.58, №6, с.6-17.
23. Ососков Г.А. В кн.: Совещание по программированию и математическим методам. Дубна, 1977. ОИЯИ, Д10-11-11264, Дубна, 1978, с.268-272.
24. Frei W., Chen C.-C. Fast Boundary Detection a Generalization and a New Algorithm. IEEE, 1977, tr.C-26, No.10, pp.988-997.
25. Обработка изображений при помощи ЦВМ. ТИИЭР, 1972, т.60, №7, с.204.
26. Вылов Ц. и др. ОИЯИ, 10-7034, Дубна, 1973.
27. Орманджиев С.И. и др. ОИЯИ, 13-8797, Дубна, 1975.
28. Соучек Б. Мини-ЭВМ. "Мир", М., 1976, с.520.
29. Гонусек М., Фромм В.О. ОИЯИ, 10-10007, Дубна, 1976.
30. Волков Н.Г., Цупко-Ситников В.М., Чураков А.К. ОИЯИ, 10-12400, Дубна, 1979.
31. Ramamoorthy C.V., Turner J.L., Wah B.W. IEEE, 1978, tr.C-27, No.9, pp.800-815.
32. Фенггенден Р.Г. и др. ПТЭ, 1968, №3, с.68.
33. Schmitt H.W. NIM, 1968, vol.63, p.237.
34. Bartsch H. et al. NIM, 1974, vol.121, pp.185-189.
35. Балагуров А.М. и др. ОИЯИ, Р14-12840, Дубна, 1979.

36. Ahmed N., Rao K. Orthogonal Transform. Springer-Verlag, 1975, p.263.
37. Федоров Г.А., Мухин В.И. В кн.: Прикладная ядерная спектроскопия, 1975, вып.5.
38. Haines E.L., Parker R.H. NIM, 1969, vol.69, No.1, p.11.
39. Dillard G.M. Proc. IEEE, 1976, Nat. Aerospace and Electron. Conf. Dayton, pp.101-105.
40. Andrew H., Patterson C. IEEE, 1977, tr. C-25, No.7, pp.141-147.
41. Pratt W.K. Slant Transform Image Coding. IEEE, 1973, tr.C-22, No.8, pp.1075-1093.
42. Whitehead A.V. et al. IEEE, 1967, tr. NS-14, No.1, p.599.
43. Сороко Л.М., Стриж Т.А. ОИЯИ, Б1-10-8364, Дубна, 1974.
44. Вокадерная телефония. Методы и проблемы /сборник статей под ред. А.А.Пирогова/. "Связь", М., 1974, с.336.
45. Баглай Р.Д., Смирнов Н.К. ЖВМ и МФ, 1975, №1, с.241.
46. Бялко А.А. и др. ОИЯИ, P10-80-107, Дубна, 1980.
47. Байла И., Ососков Г.А. ОИЯИ, P10-80-389, Дубна, 1980.
48. Байла И., Ососков Г.А., Приходько В.И. ОИЯИ, P10-80-162, Дубна, 1980.
49. Коган И.М. Прикладная теория информации. "Радио и связь", М., 1981, с.216.

Рукопись поступила в издательский отдел
26 апреля 1982 года.

Бялко А.А., Волков Н.Г., Цупко-Ситников В.М. P10-82-304
Методы сжатия данных в задачах экспериментальной ядерной физики

Проведен анализ и дана классификация методов сжатия данных, используемых в ядерной физике, в зависимости от применяемых в них подходов к оценке содержащейся в данных информации - классического /шенноновского/ или алгоритмического /предложенного А.Н.Колмогоровым/. Из трех введенных категорий методов: а/ основанных на использовании статистических закономерностей, б/ использующих элементы обработки и в/ базирующихся на ортогональных преобразованиях, наиболее эффективны по степени сжатия данных два последних. Отмечается перспективность методов, использующих алгоритмы с базисом преобразования, определяемым самими данными. Эти методы имеют высокие показатели сжатия и в то же время позволяют сохранить информацию, близкую к исходной.

Работа выполнена в Лаборатории ядерных проблем ОИЯИ.
Препринт Объединенного института ядерных исследований. Дубна 1982

Bialko A.A., Tsupko-Sitnikov V.M., Volkov N.G. P10-82-304
Data Compression Methods in Problems Experimental Nuclear Physics

The analysis and classification of the data compression methods, used in nuclear physics, based on the way for evaluation of information (classical - given by Shannon, or algorithmic - given by A.N.Kolmogorov) are presented. Among three suggested classification categories of the methods: a) ones based on the use of statistical relations; b) ones using elements of data analysis and c) ones based on orthogonal transformations - the most effective are the last two. The methods using algorithms with transformation basis, determined by the data, are especially promising. These methods have high compression degree and at the same time permit to conserve information very close to initial.

The investigation has been performed at the Laboratory of Nuclear Problems, JINR.

Preprint of the Joint Institute for Nuclear Research. Dubna 1982

Перевод О.С.Виноградовой.