

сообщения
Объединенного
института
ядерных
исследований
Дубна

3379/82

19/4-82

10-82-292

+

З.М.Иванченко, И.М.Иванченко, Н.Н.Карпенко,
П.В.Мойсенз

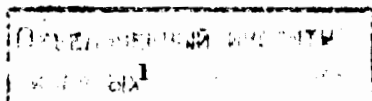
СРАВНИТЕЛЬНЫЕ ХАРАКТЕРИСТИКИ
ПРОГРАММ НАКОПЛЕНИЯ
СТАТИСТИЧЕСКИХ РАСПРЕДЕЛЕНИЙ
В СИСТЕМЕ НВООК

1982

Система нвook /1,2/ - пакет прикладных программ, предназначенный для статистической обработки данных, - обеспечивает декларирование статистических объектов, создание динамических структур локальной базы данных, накопление распределений и операции над ними, включая фитирование, сглаживание и т.п. В нвook входят программы представления результатов на АЦПУ в виде гистограмм, диаграмм рассеяния, таблиц. Для графического представления результатов, получаемых нвook, может быть использован разработанный в ЦЕРНе пакет nplot /3/. Программы пакетов написаны на подмножестве фортрана, близком к ANSI FORTRAN4. Дополнительные свойства мобильности (портабельности) обеспечиваются за счет средств генерации типа patchy /4/ или update /5/. нвook используется на ЭВМ CDC-6500, ЕС ЭВМ, БЭСМ-6 ОИЯИ и на других ЭВМ/1/ в различных ядерно-физических центрах. При выработке основных концепций, правильность которых подтверждена опытом широкого и интенсивного использования этой системы, учитывались такие требования, как надежность, многофункциональность, адаптивность и простота использования.

При обработке большого количества событий (десятки миллионов в современных электронных экспериментах) и широком множестве типов гистографируемых величин одним из наиболее критичных к экономичности пакета параметров является скорость накопления гистограмм. В рамках нвook имеются программы nfill, nfi(2), обеспечивающие существенно различные скорости накопления статистических распределений. В последнее время предприняты попытки создания новых программ, обладающих повышенной скоростью выполнения алгоритма.

Программа nfill характеризуется универсальностью и адаптивностью к различным формам представления входной информации, позволяет за счет автоматической настройки накапливать распределения различной размерности. Используя локальную базу данных, программа идентифицирует статистический объект и увеличивает содержимое канала,



соответствующего гистограммируемым величинам, на значение веса, задаваемого вызывающей программой или по умолчанию. Гистограммируемые величины могут быть как вещественными, так и целыми. Распределения строятся в упакованном виде и в режиме упаковки нескольких каналов в одном машинном слове. Ширина канала может задаваться вызывающей программой или выбираться автоматически. При дефиците оперативной памяти банк данных соответствующего статистического объекта автоматически размещается во вторичной памяти. При заполнении двумерных распределений программа HFILE накапливает также общие и парциальные проекции, если они продекларированы на стадии заказа. В программе HFILE последовательно используется концепция защитного программирования. Вначале осуществляется проверка входных данных на соответствие их свойств атрибутам и диапазонам изменения, на их логическую согласованность. При этом используются информационные блоки, поставляемые на начальном шаге — заказе статистических объектов. Защитная часть программы позволяет использовать данные, накопленные в других задачах, без явного описания типов этих данных. Таким образом, снижаются требования к формам представления гистограммируемых данных, что особенно актуально при решении крупной проблемы в рамках многостороннего сотрудничества, когда осуществляется обмен информацией на так называемых лентах суммарных результатов. Обладая рядом указанных привлекательных качеств, эта программа уступает остальным программам пакета по скорости накопления данных.

В программе HF1(2) скорость накопления повышена по сравнению с HFILE за счет специализации, сужения защитных функций и снижения адаптивности. При оперировании одномерными распределениями (значительное количество параллельно накапливаемых гистограмм, режим упаковки) программа HF1 работает более чем в 2 раза быстрее программы HFILE. Одним из возможных подходов применения этих программ является использование HFILE на ранней стадии создания системы обработки с последующим переходом на более быстрые программы накопления.

Возрастание объема данных, подлежащих статистической обработке, внедрение пакета HBOOK как в системах реального времени, так и на последующих стадиях обработки экспериментальной информации стимулировали разработку программ с повышенной скоростью накопления гистограмм. Одно из направлений этих разработок связано с внедрением ЕС ЭВМ в ОИЯИ. Первоначально программа HFF1 (быстрого накопления распределений) разрабатывалась как расширение существующего функционального набора. В процессе разработок и исследования временных характеристик новой программы получены решения, позволяющие заменить

программу HF1, используя вместо нее HFF1. Сопровождение обеих указанных программ в будущем предопределяется только требованием обратной совместимости, т.к. обращение к HFF1 несколько отличается от оператора вызова HF1 в программах пользователя.

Одним из основных требований к пакету программ HBOOK является портабельность пакета в целом, обеспечиваемая в значительной степени за счет свойств генерируемости и унификации библиотек общего назначения. Дополнительным требованием является прямая портабельность его базовых компонентов. В связи с этим созданы два варианта HFF1. Портабельный вариант реализован на стандартном подмножестве фортрана с учетом концепций структурного программирования. Выигрыш в скорости в данном варианте получен, в частности, за счет внутримодульной реализации функций сжатия результирующих массивов и уменьшения объема вычислений квазидинамических характеристик. Дополнительный выигрыш в скорости получен за счет алгоритма статификации обрабатываемых массивов, в рамках которого указатели элементов динамической структуры выносятся в программу пользователя.

Для типовых режимов статистической обработки машинно-независимый вариант HFF1 дает шестикратное увеличение скорости по сравнению с HFILE. Этот фактор возрастает с увеличением количества накапливаемых распределений.

В связи с широким внедрением ЕС ЭВМ, отражающим общую тенденцию унификации средств вычислительной техники в рамках международного сотрудничества, изменилось содержание понятия мобильности программ для ЭВМ различной конфигурации.

Вариант ЕС-ориентированной программы HFF1 написан на языке ассемблера. Увеличение скорости в данном варианте по сравнению с фортраном получено за счет использования регистровых операций, оптимизации алгоритмов вычисления адресов динамических структур на основе входных данных смешанного типа. В режиме сжатия результирующих массивов основной выигрыш в скорости достигается путем учета байтовой организации памяти ЭВМ. Оптимизация скорости обеспечивается также за счет векторизованных (групповых) операций пересылки значений переменных и линейаризованного алгоритма передачи фактических параметров. Общий фактор повышения скорости ЕС-ориентированного варианта HFF1 по сравнению с HFILE достигает 10 в типовых режимах, а для реализации процедуры упаковки необходимо в 13 раз меньше времени.

Следует отметить относительно сильную зависимость времени выполнения программ от уровня транслятора на ЕС-ЭВМ. Мы использовали при создании объектных модулей транслятор типа H.

Временные характеристики программ накопления

ЭВМ	Програм- ма	Время накопления в относительных единицах					
		без упаковки			с упаковкой		
		I	IO	IOO	I	IO	IOO
EC-1040	HFILL	15,29	19,23	21,53	17,73	20,49	22,43
	HF1	3,22	7,39	9,52	7,65	10,26	12,13
	HFF1 (A)	2,17	2,36	2,36	2,21	2,44	2,44
EC-1060	HFILL	6,90	8,60	9,77	7,82	9,21	10,21
	HF1	1,37	2,95	3,90	3,20	4,14	4,96
	HFF1 (F)	1,42	1,33	1,33	2,45	2,27	2,27
	HFF1 (A)	I	I	I	I	I	I
CDC-6500	HFILL	6,98	7,60	8,57	7,61	8,00	8,83
	HF1	1,14	2,70	3,67	2,86	3,81	4,60
	HFF1 (F)	1,20	1,54	1,54	2,47	2,49	2,49
CYBER 170-720	HFILL	6,63	7,09	7,97	7,15	7,39	8,11
	HF1	1,03	2,45	3,30	2,51	3,38	4,12
CYBER 170-730	HFILL	4,27	4,65	5,22	4,71	4,86	5,34
	HF1	0,73	1,68	2,25	1,69	2,29	2,79
CDC-7600	HFILL	0,57	0,61	0,67	0,63	0,64	0,69
	HF1	0,11	0,20	0,26	0,23	0,29	0,34

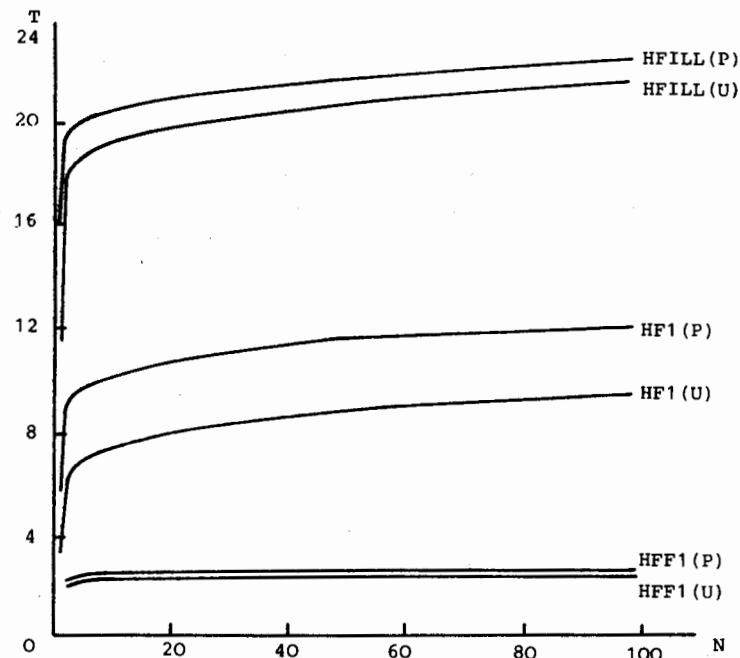
Обращение к программе HFF1

CALL HFF1 (ID, NID, X, WEIGHT)

обеспечивает увеличение содержимого канала гистограммы, соответствующего значению x , на величину WEIGHT.

ID - идентификатор гистограммы (входной параметр, имеющий тот же смысл, что и в других программах пакета /I/).

NID - выходной параметр. Этой переменной перед первой операцией заполнения гистограмм необходимо присвоить отрицательное значение на тех ЭВМ, где оператор ветвления не допускает обработки неопределенных величин. Между ID и NID должно быть взаимно однозначное соответствие.



Временные характеристики программ накопления для ЭВМ EC-1040.

- N - число гистограмм,
- T - время накопления в относительных единицах,
- (P) - с упаковкой содержимого каналов,
- (U) - без упаковки содержимого каналов.

Ассемблерный вариант программы HFF1 допускает три уровня квантования (1 байт, 2 байта, полное слово) при размещении содержимого каналов гистограммы в памяти ЭВМ. Нарушение этого условия идентифицируется и диагностируется программой.

В ОИЯИ на CDC-6500 внедрен фортранный вариант HFF1, а на EC ЭВМ - ассемблерный (в таблице они помечены как HFF1 (F) и HFF1 (A) соответственно).

Синтаксическая и функциональная совместимость программ HF1 и HFF1 позволяет достаточно просто изменить ранее созданные программы пользователя с целью улучшения их временных характеристик.

Для выбора программы накопления, а также режима работы выбранной программы для конкретных применений могут быть использованы временные характеристики программ, приведенные в таблице и на рисунке. Эти данные позволяют также оценить вычислительную мощность раз-

личных ЭВМ относительно базовой для широкого класса программ обработки.

В заключение авторы выражают благодарность профессору Н.Н.Говоруну за поддержку данной работы, И.А.Савину за инициирование работ по созданию ЕС-ориентированной версии программы.

Мы признательны Э.Штрайту за помощь при анализе влияния различного уровня фортранных трансляторов на эффективность программ в машинном коде; В.Н.Евсиной за помощь на стадии подготовки и отладки программ на ЕС-1060.

Авторы благодарят Р.Бруна за полезные обсуждения алгоритмов оптимизации доступа к элементам динамической структуры.

Литература

1. Brun R., Ivanchenko I. Palazzi P. HBOOK-3.00, CERN, DD/77/9, 1977. HBOOK-3.18, CERN, DD-EE-81.1, 1981.
2. Брун Р. и др. ОИЯИ, Р10-11850, Дубна, 1978.
3. Brun R., Watkins H. HPLOT-4, CERN, DD-EE-80.2.
4. Klein H., Zoli J., PATCHY-4, CERN, 1977.
5. UPDATE Reference Manual. CDC, Publ. No. 60342500, 1974.

Рукопись поступила в издательский отдел
19 апреля 1982 года.