

докл. VI Всесоюз. школы-семинара (Львов, 1987 г.). — Львов : Физико-мех. ин-т им. Г. В. Карпенко АН УССР, 1987. — Ч. 1. — С. 140—141.

5. Сигорский П. В. Методика построения и расчета интегрированных систем обработки информации // Там же. — С. 189—190.

6. Сигорский П. В. Статистические характеристики интегрированной системы обработки океанографических данных // Вычисл. и прикл. математика. — 1987. — Вып. 62. — С. 113—122.

7. Сигорский П. В. Аналитическое моделирование систем интеграции пакетов прикладных программ // Вопр. кибернетики. — 1987. — Вып. 123. — С. 151—161.

Б-НО

Академия наук Украинской ССР
Ордена Ленина Институт кибернетики имени В. М. Глушкова

На правах рукописи

СИГОРСКИЙ Петр Витальевич

УДК 681.3.06

**МЕТОД ПОСТРОЕНИЯ ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ
АВТОМАТИЗИРОВАННЫХ СИСТЕМ ОБРАБОТКИ
ОКЕАНОГРАФИЧЕСКИХ ДАННЫХ**

05.13.11 — математическое и программное обеспечение вычислительных машин и систем

Автореферат диссертации на соискание ученой степени
кандидата физико-математических наук

Подп. в печ. 21.08.87. БФ 11462. Формат 60×84/16. Бум. писч. Офс. печ.
Усл. печ. л. 0,93. Усл. кр.-отг. 0,93. Уч.-изд. л. 1,0. Тираж 100 экз. Заказ 1705. Бесплатно.

РИО Института кибернетики имени В. М. Глушкова АН УССР
252207 Киев 207, проспект Академика Глушкова, 20

Киев — 1987

K

Работа выполнена в ордена Ленина Институте кибернетики имени В. М. Глушкова АН УССР.

Научный руководитель: кандидат технических наук, старший научный сотрудник
ПЛАТОНОВ Б. А.

Официальные оппоненты: доктор физико-математических наук, профессор ВЕЛЬБИЦКИЙ И. В.,
кандидат физико-математических наук ИСАЕВ И. В.

Защита состоится «16» октября 1987 г. в 14 часов на заседании специализированного совета К 016.45.01 при Институте кибернетики имени В. М. Глушкова АН УССР по адресу:
252207 Киев 20

С диссертационным архивом и

техничес-

Автореферат

Ученый специалист

Г. А.

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы. Успешное освоение Мирового океана тесно связано с анализом и обработкой океанологической информации. Средством эффективного решения указанных задач стало создание мощных информационно-вычислительных систем для автоматизации обработки океанографических данных. Однако на пути решения стоящих перед такими системами задач встречается ряд трудностей, основными из которых являются следующие:

пространственно-временная дискретность фактически измеренных океанографических данных и связанная с этим сложность процедур обработки, необходимых для восстановления интересующих значений переменных в любой произвольной точке океана, а также оценка степени их достоверности;

большая номенклатура данных и значительный общий объем информации, разнообразие структур и форматов ее представления;

большое количество частных схем обработки и многообразие программного обеспечения;

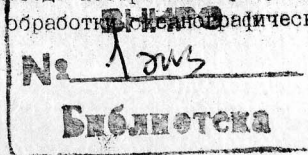
сложность задач анализа океанографических данных;

большое число центров обработки океанографической информации и фактическое отсутствие в них унифицированного программного и информационного обеспечения.

Кроме того, существующее программное обеспечение для обработки океанографических данных создавалось, в основном, для решения узкоспециализированных классов задач. При решении более широкого класса задач такое программное обеспечение не всегда оказывается достаточно эффективным. Повышение эффективности систем обработки океанографических данных должно строиться на основе системного подхода, базирующегося на концепциях информационного единства теории и эксперимента, а также оптимизации информационно-вычислительных систем.

Все вышесказанное позволяет выделить задачу анализа и обработки океанографических данных в отдельный класс задач во всем множестве задач анализа и обработки данных на ЭВМ и обуславливает необходимость проведения исследований, направленных на создание новых универсальных методов построения различного рода систем обработки океанографической информации, отвечающих современному уровню развития вычислительной техники.

Цель диссертационной работы - исследование и разработка принципов интеграции пакетов программ различного назначения и создание на их основе метода построения программного обеспечения автоматизированных систем обработки океанографических данных.



Научная новизна и практическая ценность. Введено понятие интегрированной системы пакетов программ обработки данных. В рамках этого определения дана корректная постановка задачи выбора наилучшего, в некотором смысле, алгоритма обработки данных из имеющегося в интегрированной системе набора алгоритмов.

Для решения задач интеграции пакетов программ обработки данных предложен подход, основанный на алгебраической теории коррекции эвристических процедур анализа данных, созданной в БЦ АН СССР под руководством Ю.И. Журавлева. В рамках этого подхода в диссертации разработан метод построения программного обеспечения автоматизированных систем обработки данных, который был развит для области океанографических данных.

Описан способ синтеза корректного алгоритма выбора в полной системе базисных операторов распознающего оператора в форме, удобной для реализации на ЭВМ. Обоснована эффективность этого способа для случая, когда исходная совокупность моделей состоит из эвристических алгоритмов, линейных моделей, L -моделей и моделей вычисления оценок.

Для расчета характеристик интегрированных систем в диссертации использованы методы аналитического моделирования.

Разработанный в диссертации метод построения программного обеспечения использовался при создании Автоматизированной системы обработки океанографических данных (АСОД) (тема 0.74.01.08.01.02 - Постановление КНТ СССР № 475/251/131 от 26.12.1980 г., дополнительная тема 0.74.01.08.01.07 - Постановление КНТ СССР № 353 от 15.07.1983 г. и Президиума АН УССР № 1453 от 9.08.1983 г.), которая внедрена и успешно функционирует в ряде мореведческих организаций страны.

Апробация работы. Основные положения и результаты диссертации докладывались и обсуждались на Всесоюзной научно-технической конференции "Моделирование-85. Теория, средства, применение" (Киев, 1985 г.) на Одиннадцатом Всесоюзном семинаре по вычислительным сетям (Рига, 1986 г.), на Всесоюзной конференции "Автоматизированные системы обработки изображений" (Львов, 1986 г.), на VI Всесоюзной школе-семинаре "Распараллеливание обработки информации" (Ивано-Франковск, 1987 г.), на заседании республиканского семинара "Операционные системы и связь человека с ЭВМ" научного совета АН УССР по проблеме "Кибернетика" (Киев, 1987 г.), на заседаниях семинара лаборатории № 421 Института кибернетики имени В.М. Глушкова АН УССР (Киев, 1982-1986 гг.).

Публикации. По теме диссертации опубликовано семь печатных работ, список которых приведен в конце автореферата.

Структура и объем работы. Диссертационная работа изложена на 98 страницах машинописного текста и состоит из введения, четырех глав, заключения и библиографии. Список литературы насчитывает 72 наименования. Представлено 9 рисунков.

СОДЕРЖАНИЕ РАБОТЫ

Во введении обосновывается актуальность темы, приводятся обзор и анализ современного состояния проблемы обработки океанографической информации, обзор результатов в области алгебраической теории распознавания, дается краткая характеристика основных положений диссертации, ее содержание и структура.

В первой главе даются основные определения и постановка задачи выбора наилучшего алгоритма обработки данных из имеющегося в интегрированной системе набора алгоритмов. Указанная задача сформулирована как задача распознавания.

Практика обработки океанографической информации показывает, что в настоящее время для анализа каждого вида океанографического запроса разработано большое количество прикладных программ, часто реализующих один и тот же или близкие виды обработки данных. Это обусловлено, с одной стороны, тем, что эти программы написаны различными коллективами авторов, а, с другой стороны, тем, что специфика форматов исходной информации в значительной степени определяет алгоритм анализа и обработки данных. В связи с этим актуальной является постановка и решение следующей задачи: как по виду исходной информации указать оптимальный алгоритм ее обработки из имеющегося в системе набора алгоритмов.

Для корректной постановки этой задачи необходимо дать определение пакета программ, обладающее достаточной для этого математической строгостью. В диссертации показано, что сформулированная задача может рассматриваться как задача распознавания и, следовательно, определение пакета программ должно даваться в терминах, адекватных понятиям теории распознавания. В настоящее время наиболее распространенным в задачах распознавания является алгебраический подход, разработанный в БЦ АН СССР под руководством Ю.И. Журавлева. В рамках используемых при таком подходе понятий оказалось возможным дать строгое определение пакета программ, адекватное рассмотренным в диссертации задачам. При этом пакет программ есть некоторая структура, составленная из элементарных объектов или образующих. Для наших целей в

качестве образующих выбираются элементарные виды обработки океано-графической информации. Каждому запросу на обработку данных соответствует некоторая комбинация образующих.

С каждой образующей связаны два типа свойств. Первый - это признак вида обработки данных. Второй тип свойств определяется множеством типов данных, связанных с образующей.

Образующую можно рассматривать двумя способами. В первом случае она рассматривается в абстрактном виде без учета среды, в которой она функционирует. В этом случае она является отвлеченным символом и не представляет интереса для нашего исследования. Противоположный случай, рассматриваемый ниже, состоит в том, что каждая образующая погружается прямо или опосредованно через другие образующие в информационную среду. Информационная среда есть совокупность двух множеств - множества начальных J_0 и множества финальных J_f информационных. Каждое из этих множеств есть объединение некоторых типов данных.

Множество образующих G состоит из непересекающихся классов образующих $G^\alpha, G^\alpha \subset G$, где α - индекс класса образующих.

Под пакетом программ будем понимать структурное объединение стандартных блоков (т.е. образующих) в конфигурацию.

Для выделения регулярных или допустимых конфигураций рассмотрим систему ограничений R связанных только с согласованием типов данных образующих. А именно, будем говорить, что структура связей $S \subset G \times G$ является допустимой, если для каждой связи $g_i (\beta_{k_1}, \beta_{k_2}) g_2, g_1, g_2 \in G, \beta_{k_1} \in \Omega_{out}(g_1)$ - выходная, $\beta_{k_2} \in \Omega_{in}(g_2)$ - входная связи образующих g_1 и g_2 соответственно и $T_{k_1} \subseteq T_{k_2}$, где $T_{k_i}, i=1,2$, - тип данных, соответствующий связям β_{k_i} . Множество регулярных конфигураций, получаемых с помощью множества R , будем обозначать через $\mathcal{B}(R)$. Множество $\mathcal{B}(R)$ характеризует регулярность структур.

Резюмируя вышесказанное, введем следующие определения.

Определение 1.1. Допустимая конфигурация $q \in \mathcal{B}(R)$ есть пара $q = (\hat{G}, S)$, где $\hat{G} = \{g_1, g_2, \dots, g_n\}$ - список всех образующих, входящих в q , S - допустимая структура связей.

Определение 1.2. Пакетом программ называется допустимая конфигурация $q = (\hat{G}, S)$.

Перейдем к рассмотрению вопросов, связанных с композицией допустимых конфигураций.

Пусть имеется два пакета программ (две допустимые конфигурации) $q_1 = (\hat{G}_1, S_1), q_2 = (\hat{G}_2, S_2)$, где $\hat{G}_i, S_i, i=1,2$ - списки образующих

и структур связи соответственно. Определим композицию допустимых конфигураций q_1 и q_2 как конфигурацию $q = (\hat{G}, S)$, где $\hat{G} = \hat{G}_1 \cup \hat{G}_2$, а S определяется следующим образом: пусть $T_1'', T_2'', \dots, T_{k_2}''$ - входные типы данных второго пакета, а $T_1', T_2', \dots, T_{k_1}'$ - выходные типы данных первого; тогда для тех i, j , для которых $T_i'' \subseteq T_j'$, устанавливается связь между соответствующими образующими, а остальные получают значения исходных или выходных наборов данных в зависимости от принадлежности к первому или второму пакету. Пусть теперь $\Omega(q_1)$ и $\Omega(q_2)$ - множества, образованные внешними связями конфигураций q_1 и q_2 соответственно, а S_{12} представляет собой список соединений связей, принадлежащих $\Omega(q_1)$, со связями, принадлежащими $\Omega(q_2)$, при условии, что устанавливаются только попарные соединения и, следовательно, групповые соединения отсутствуют. Тогда из определения композиции допустимых конфигураций следует, что $q \in \mathcal{B}(R)$ в том и только в том случае, если $S \in \Sigma$ и $\beta \rho \beta'$ выполняется для всех новых связей, соединенных в соответствии с S_{12} , Σ - тип соединений конфигураций в рассматриваемом множестве регулярных конфигураций $\mathcal{B}(R)$, ρ - отношение согласования.

Композиция большего чем два числа пакетов программ определяется как последовательная композиция

$$q_1 \circ q_2 \circ \dots \circ q_n = (\dots ((q_1 \circ q_2) \circ q_3) \circ \dots) \circ q_n.$$

В случае композиции нескольких допустимых конфигураций в результирующей конфигурации могут появиться подобные образующие, т.е. те, которые имеют равные входные и выходные арности и реализуют один и тот же вид обработки информации, но разными способами. Это связано с тем, что в различных пакетах программ для обработки одного и того же запроса могут быть использованы отличные друг от друга алгоритмы, что обусловлено, в частности, особенностями исходной информации.

Путем в допустимой конфигурации $q \in \mathcal{B}(R)$ будем называть последовательность образующих g_1, g_2, \dots, g_k таких, что g_1 имеет на входе по крайней мере один набор данных, g_k имеет на выходе хотя бы один выходной набор данных и для всех $2 \leq i \leq k$ образующие g_{i-1} и g_i связаны между собой.

Поскольку в конфигурации, представляющей собой композицию допустимых конфигураций, могут быть подобные образующие, то на множестве всех образующих этой конфигурации может быть определено разбиение на классы эквивалентности, где в один класс отнесены все подобные образующие.

Каждая допустимая конфигурация образует структуру, т.е. на множестве всех ее образующих \hat{G} может быть построено транзитивное

отношение порядка. В самом деле, на множестве образующих \mathcal{G} может быть построено разбиение на слои. В первый слой относятся те элементы g_i , которые в качестве входных связей имеют хотя бы один исходный набор данных, в j -й слой относятся те элементы множества \mathcal{G} , которые имеют хотя бы одну входную связь, инцидентную элементу слоя $(j-1)$, и не имеют связей с элементами слоев меньшего порядка.

Вышесказанное позволяет определить каноническую структуру записи композиции допустимых конфигураций следующим образом.

Пусть есть конфигурация $q = (\mathcal{G}, S)$, являющаяся композицией допустимых конфигураций q_1, q_2, \dots, q_m и пусть на множестве \mathcal{G} построено транзитивное отношение порядка. Тогда, поскольку в каждом слое множества \mathcal{G} могут быть подобие образующие, на множестве образующих каждого слоя может быть определено разбиение на классы эквивалентности, где в один класс отнесены все подобные образующие фиксированного слоя множества \mathcal{G} . Таким образом вычисление любого информационного запроса может быть выполнено не более чем за ℓ шагов, где ℓ - число слоев в канонической записи допустимой конфигурации, соответствующей некоторой композиции пакетов программ.

В силу того что на каждом слое определены классы эквивалентных образующих, необходимо уметь из числа всех эквивалентных цепочек выбирать в некотором смысле оптимальную цепочку для вычисления фиксированного запроса на обработку океанографической информации. Эквивалентными считаются цепочки, состоящие из попарно подобных образующих. В диссертации указанная задача рассматривается как задача распознавания. В общем виде она может быть представлена следующим образом. Будем считать, что пакет программ $q = (\mathcal{G}, S)$ вычисляет запросы z_1, z_2, \dots, z_n и для каждого запроса z_i в пакете имеется K_i эквивалентных между собой путей вычисления. Через $M = (M_1, M_2, \dots, M_n)$ обозначим эталонную информацию для вычисления запросов z_1, z_2, \dots, z_n . Каждое M_i состоит из нескольких наборов исходной информации, для каждого из которых указан оптимальный путь вычисления в конфигурации q . На основе этой информации для вновь поступившего запроса и соответствующей ему информации необходимо выбрать наилучший путь решения.

В диссертации эта задача решается при помощи n -годов алгебраической теории коррекции эвристических процедур.

Во второй главе приведено исследование пространства эвристических процедур с целью получения корректных алгоритмов решения задачи выбора оптимального пути в конфигурации. Описан способ синтеза корректного алгоритма выбора в полной системе базисных опе-

раторов распознающего оператора в форме, удобной для реализации на ЭВМ.

Пусть задано непустое множество \mathcal{O} допустимых объектов. Предполагается, что на всем множестве допустимых объектов определено разбиение на конечное число классов K_1, K_2, \dots, K_ℓ :

$$\mathcal{O} = K_1 \cup K_2 \cup \dots \cup K_\ell.$$

Вектор $\tilde{\alpha}(S) = (\alpha_1(S), \dots, \alpha_\ell(S))$, где $\alpha_i(S)$ - значение предиката

$$P_j(S) \equiv "S \in K_j", \quad j = 1, 2, \dots, \ell,$$

назовем информационным вектором объекта S по классам K_1, \dots, K_ℓ . Множество

$$I_0 = \{S_1, \dots, S_m, \tilde{\alpha}(S_1), \dots, \tilde{\alpha}(S_m)\}$$

называется стандартной обучающей информацией, а множество $I = \{S_1, S_2, \dots, S_m\}$ - обучающим множеством. Множество $\tilde{S}^2 = \{S^1, S^2, \dots, S^2\}$, такое, что $S^j \notin I, j = 1, 2, \dots, \ell$, называется контрольным множеством. Пара (I_0, \tilde{S}^2) называется стандартной исходной информацией для задачи распознавания Z .

Задача распознавания Z , заданная в стандартной форме, определяется парой (I_0, \tilde{S}^2) . Положим $P_j(S^i) \equiv "S^i \in K_j" = \beta_{ij} \in \{0, 1\}, i = 1, 2, \dots, \ell; j = 1, 2, \dots, \ell$. Задача распознавания состоит в том, чтобы указать алгоритм A такой, что

$$A(Z) = A(I_0, \tilde{S}^2) = \|\beta_{ij}\|_{\ell \times \ell}.$$

При этом если $\beta_{ij} = \beta_{ij}^*, i = 1, 2, \dots, \ell; j = 1, 2, \dots, \ell$, то алгоритм называется корректным для задачи Z .

Первые работы по построению корректных алгоритмов были проведены в рамках алгебраических замыканий фиксированных моделей. Так, в фундаментальной работе Ю.И. Журавлева^х приводится построение такого алгоритма для модели алгоритма вычисления оценок. Однако приведенный там алгоритм, оставаясь плодотворным в теоретических исследованиях, является неприемлемым для практического использования. Дело в том, что только описание такого алгоритма требует, в общем случае, расхода $const \cdot \ell^2$ ячеек памяти. Для преодоления этой трудности предлагается другой подход, который использует общую форму записи корректного алгоритма из работы Ю.И. Журавлева, однако слагаемые ре-

^х Журавлев Ю.И. Об алгебраическом подходе к решению задач распознавания или классификации // Пробл. кибернетики. - 1976. - Вып. 33. - С. 5-68.

зультрирующего полинома строится другим образом.

Пусть задано семейство $\{A\}$ алгоритмов для решения задач распознавания. Каждый такой алгоритм A может быть представлен в виде $A = B \cdot C$, где B - распознающий оператор, т.е.

$$B(I_0, \tilde{S}^i) = \|\Gamma_{ij}\|_{q,e},$$

Γ_{ij} - действительные числа, C - фиксированное пороговое решающее правило.

На множестве $\{B\}$ распознающих операторов введем операции сложения, умножения и умножения на скаляр. Пусть $B', B'' \in \{B\}$, b - скаляр:

$$B'(I_0, \tilde{S}^i) = \|\Gamma'_{ij}\|_{q,e}, \quad B''(I_0, \tilde{S}^i) = \|\Gamma''_{ij}\|_{q,e}.$$

Определим операторы $b \cdot B'$, $B' \cdot B''$, $B' \cdot B''$ согласно правилам

$$b \cdot B'(I_0, \tilde{S}^i) = \|b \Gamma'_{ij}\|_{q,e}, \quad (1)$$

$$B'(I_0, \tilde{S}^i) \cdot B''(I_0, \tilde{S}^i) = \|\Gamma'_{ij} \cdot \Gamma''_{ij}\|_{q,e}. \quad (2)$$

$$B'(I_0, \tilde{S}^i) \cdot B''(I_0, \tilde{S}^i) = \|\Gamma'_{ij} \cdot \Gamma''_{ij}\|_{q,e}. \quad (3)$$

Замыкание множества $\{B\}$ относительно операций (1), (2) называется линейным замыканием $\mathcal{L}(B)$, а относительно операций (1), (2), (3) - алгебраическим замыканием $\mathcal{A}(B)$.

Множество пар индексов $\{(i, j) | i=1, 2, \dots, j; j=1, 2, \dots, e\} = \hat{M}_{qe}$ информационной матрицы $\|\beta_{ij}\|_{q,e}$ представим в виде

$$\hat{M}_{qe} = M_{qe}^0 \cup M_{qe}^1, \quad M_{qe}^1 = \{(u, v) | \beta_{uv} = 1\}, \quad i \in \{0, 1\}.$$

Эвристический алгоритм A (соответствующий ему оператор B) называется допустимым для задачи $Z(I_0, \tilde{S}^i)$, если $B(Z) = \|\Gamma_{ij}\|_{q,e}$ и $\max_{i,j} \|\Gamma_{ij}\|_{q,e} = \Gamma_{zt} > 0$, $(z, t) \in M_{qe}^1$.

Определение 2.4. Точка $(u, v) \in M_{qe}^1$ называется отмеченной допустимым оператором B (алгоритмом A), если $\Gamma_{uv} > \Gamma_{zw}$ для всех $(z, w) \in M_{qe}^1$.

Пусть в исходном семействе операторов $\{B\}$ выделена конечная совокупность операторов $\{B_1, B_2, \dots, B_t\}$, допустимых для задачи Z и таких, что

$$\bigcup_{i=1}^t M_{qe}(B_i) = M_{qe}^1. \quad (4)$$

Тогда оператор

$$\tilde{B} = (c_1 + c_2) \sum_{i=1}^t \tilde{B}_i^{k_2},$$

$$k_2 = \{(ent + en(c_1, c_2) - |ca c_1|) / |ca B(z)|^{-1}\} \cdot t,$$

$$\tilde{B}_i = \frac{1}{a(z)} B_i(Z), \quad a(z) = \min_j \Gamma_{ij}^2,$$

является корректным для задачи Z . Множество $\{B_1, B_2, \dots, B_t\}$, удовлетворяющее (4), называется базисной системой распознающих операторов.

Из вышесказанного видно, что для построения корректного алгоритма с пороговым решающим правилом достаточно найти базисную систему распознающих операторов, такую, что отмеченные этими операторами множества покрывают множество пар индексов. Этот результат используется при построении базисной системы распознающих операторов.

Найдем множество пар (u, v) , отмеченных оператором ближайшего соседа и среднего расстояния. Определим расстояние от объекта до множества следующим образом:

$$\rho_M(S, M) = \min_{t \in M} \rho(S, t).$$

Зафиксируем $S^i \in \bar{K}^j = \tilde{S}^i \cap K_j$. Тогда если

$$\rho_M(S^i, \bar{K}_j) < \rho_M(S^i, \bar{K}_u), \quad u=1, \dots, j-1, j+1, \dots, q, \quad (5)$$

где $\bar{K}_i = K_i \cap I$, $i=1, 2, \dots, e$, то оператор ближайшего соседа

$$B_L(I_0, \tilde{S}^i) = \|\Gamma_{ij}\|_{q,e},$$

$$\Gamma_{ij} = [\rho(S^i, \bar{K}_j) + 1]^{-1}, \quad \Gamma_{ik} = 0, \quad k=1, \dots, j-1, j+1, \dots, e,$$

если (5) неверно, то $\Gamma_{iv} = 0$, $v=1, \dots, e$.

Аналогичным образом исследован оператор среднего расстояния.

Далее описаны четыре модели алгоритмов распознавания, в каждой из которых оценка $\Gamma_{ij} = \Gamma_{ij}(w_1, \dots, w_p)$ является линейной комбинацией параметров w_1, \dots, w_p . Такие модели будем называть линейными. Описаны также модели, основанные на принципе разделения.

Рассмотрим совокупность распознающих операторов

$$\tilde{B} = \{B_1\} \cup \{B_2\} \cup \{B_3\} \cup \{B_4\},$$

где $\{B_1\}$ - совокупность простых непараметрических распознающих операторов; $\{B_2\}$ - совокупность линейных моделей; $\{B_3\}$ - модели вычисления оценок; $\{B_4\}$ - Γ - модели.

Для выделения в \tilde{B} базисной системы распознающих операторов \bar{B} предлагается следующий алгоритм.

В множество \bar{B} включаем совокупность непараметрических опе-

раторов $\{B_i\}$. Находим множество $\{(u, v)\}$ пар, отмеченных операторами из $\{B_i\}$, и исключаем их из множества M_{qe}^1 . Оставшееся после этого множество пар обозначим M_{qe}^0 . Очевидно, в общем случае $M_{qe}^0 \neq \emptyset$.

Зафиксируем некоторую модель $B^i \in \{B_2\} \cup \{B_4\}$. Выберем в этой модели некоторый алгоритм B_*^i . Исключим из M_{qe}^0 все пары, отмеченные оператором B_*^i . Будем повторять выполнение этого шага до тех пор, пока в выбранной модели будут существовать операторы, отмечающие непустое множество еще не отмеченных точек множества M_{qe}^0 . Если в модели такого оператора не существует, выбирается другая модель из множества $\{B_2\} \cup \{B_4\}$ и для нее выполняются аналогичные действия.

В результате будет построена последовательность

$$\tilde{M}_1 \supset \tilde{M}_2 \supset \dots \supset \tilde{M}_r,$$

где каждому $\tilde{M}_i, i=1, \dots, r$, соответствует распознающий оператор в B^i . Напомним, что $M_{qe}(B^i)$ - множество пар, отмеченных оператором B_*^i .

Из множества $\hat{B}_2 = \{B^1, B^2, \dots, B^r\}$ выделим минимальное подмножество B_2^* , отмечающее все пары $M_{qe}(B_2)$, где

$$M_{qe}(B_2) = \bigcup_{i=1}^r M_{qe}(B^i).$$

Заметим, что по построению множества B_2 справедливо

$$M_{qe}(B^{i+1}) \setminus M_{qe}(B^i) \neq \emptyset, i=1, \dots, r-1.$$

Но тогда для построения множества B_2^* достаточно просматривать множество B_2 в порядке убывания индексов до тех пор, пока множество отмеченных этими операторами пар не станет равным $M_{qe}(B_2)$. Типичность построенного множества очевидна.

Для каждой пары из $M_{qe}^0 \setminus M_{qe}(B_2^*)$ построим оператор вычисления оценок, который ее отмечает. Множество построенных таким образом операторов обозначим через $\{B_3^*\}$.

Тогда справедлива следующая

Теорема 2.2. $\bar{B} = \{B_1\} \cup \{B_2^*\} \cup \{B_3^*\}$ - базисная система распознающих операторов.

Таким образом, построенный только что алгоритм описывает метод выбора оптимального пути в допустимой конфигурации, т.е. решает задачу выбора оптимального алгоритма обработки океанографического запроса из имеющегося в интегрированной системе набора алгоритмов.

Третья глава посвящена расчету основных характеристик оптимальных для фиксированных запросов путей в конфигурации. Этот расчет производится при помощи методов аналитического моделирования.

Функционирование интегрированных систем обработки океанографических данных может быть адекватно описано сетью пакетов программы (СПП), входящих в систему, имеющей вид направленного графа с вершинами - узлами вычисления и дугами - траекториями следования запросов.

В основе рассматриваемых в главе приближенных методов расчета положены следующие допущения:

- рекуррентность входного потока информационных запросов;
- взаимная независимость информационных потоков сети;
- достаточность среднего значения и дисперсии интервалов следования запросов в информационном потоке для его определения.

Эти допущения эквивалентны возможности декомпозиции сети на отдельные СПП с учетом замены информационных потоков СПП рекуррентным информационным потоком с теми же значениями двух первых моментов. Вносимая при этом погрешность в большинстве случаев оказывается пренебрежимо малой.

На первом этапе предполагается, что информационный поток состоит из однородных запросов ($Q=1$). При этом для определения основных характеристик интегрированных систем используются методы диффузионной аппроксимации, двумерной аппроксимации и декомпозиции систем на уровне двух моментов распределения. Затем полученные результаты обобщаются для случая, когда в СПП циркулируют запросы разных классов ($Q>1$).

Вводится понятие обобщенного запроса путем усреднения параметров вычисления и траекторий следования информационных запросов по всем классам с учетом весовых коэффициентов. Поток обобщенных запросов будет создавать эквивалентную нагрузку СПП по отношению к исходному неоднородному информационному потоку с интенсивностью

$$\lambda_i = \sum_{q=1}^Q \lambda_i^q$$

и средним временем вычисления

$M\tau_{jmi} = \sum_{q=1}^Q M\tau_{jmi}^q \lambda_i^q / \lambda_i$, где $M\tau_{jmi}^q = 1/\lambda_{jmi}^q, i=1, \dots, K$, K - число СПП в системе. Показано, что дисперсия времени вычисления обобщенного запроса находится из выражения

$$D_{jmi} = \sum_{q=1}^Q \lambda_i^q [D_{jmi}^q + (M\tau_{jmi}^q - M\tau_{jmi})^2] / \lambda_i,$$

в котором учитываются случайные колебания времени вычисления обобщенного запроса. Здесь D_{μ}^{2l} - дисперсия вычисления q -го класса запросов i -м пакетом программ.

Если n_i - среднее число обобщенных запросов в очереди к i -му ПП, то, учитывая вероятности нахождения в этой очереди запросов q -го класса, можно найти среднее число запросов q -го класса в этой очереди $n_i^q = n_i \lambda_i^q / \lambda_i$, $q = 1, 2, \dots, a$.

Рассмотренный выше подход представляется достаточно перспективным для расчета основных характеристик интегрированных систем обработки океанографической информации.

В четвертой главе приведено описание автоматизированной системы обработки океанографических данных (АСООД), построенной на принципах интеграции пакетов программ различного назначения. Как программное изделие АСООД представляет собой инструментальную систему, включающую программы предварительной обработки данных, ряд специализированных и неспециализированных пакетов программ и программных интерфейсов между ними и функционирующую в среде ОС ЕС. В качестве исходных пакетов программ выбраны пакеты РАОД, "Океанографическая станция", "Статистика", "Карта", программы-конверторы. АСООД организована таким образом, чтобы в ее состав можно было включать новые ПП, осуществляя переструктуризацию данных для работы нового пакета либо с помощью конверторов, либо посредством пакета РАОД.

Задачи, решаемые АСООД, формируются посредством специальных запросов на входном языке системы. Все пакеты в АСООД согласованы по управлению и по данным. Связь по данным осуществляется, в основном, средствами ПП РАОД. Связь по управлению между компонентами АСООД реализуется с помощью языка управления заданиями ОС ЕС.

В состав информационной базы АСООД включены данные глубоководных батометрических наблюдений - порейсово и поквадратно, наблюдения за течениями, батитермографные наблюдения, оперативные промышленные данные по морской биологии и информационно-справочная информация по морской гравиметрии.

В диссертации приведены основные технологические схемы для выполнения запросов, которые определяют набор используемых в системе пакетов программ и порядок их выполнения. Описаны основные принципы работы каждого пакета и их взаимосвязи между собой. Приведено также описание процесса решения задачи в АСООД.

В заключении перечислены основные результаты работы.

Приведен обзор и анализ современного состояния проблемы обработки океанографической информации, обзор результатов в области ал-

гебраической теории распознавания.

Дано строгое определение пакета программ, введено понятие интегрированной системы пакетов программ обработки данных. В рамках этих определений приведена корректная постановка задачи выбора наилучшего алгоритма обработки данных из имеющегося в интегрированной системе набора алгоритмов.

Предложен подход для решения задач интеграции пакетов программ, основанный на алгебраической теории коррекции эвристических процедур анализа данных. В рамках этого подхода разработан метод построения программного обеспечения автоматизированных систем обработки океанографических данных.

Построен алгоритм синтеза корректного алгоритма распознавания в форме, удобной для реализации на ЭВМ. Такой алгоритм строится как полином в алгебраическом замыкании непараметрических алгоритмов распознавания, линейных моделей, Г-моделей и моделей вычисления оценок. Обоснована эффективность этого алгоритма.

Показано, что для расчета основных характеристик интегрированных систем пакетов программ можно эффективно применять методы аналитического моделирования.

Приведено описание автоматизированной системы обработки океанографических данных (АСООД), построенной на основе интеграции пакетов программ различного назначения, которая внедрена и успешно функционирует в ряде мореведческих организаций страны.

Основные положения и результаты диссертации опубликованы в следующих работах.

1. Задорожный В. В., Сигорский П. В. Моделирование обучающей совокупности в задачах распознавания // Моделирование-85. Теория, средства, применение: Тез. докл. Всесоюз. научн.-техн. конф. (Киев, 1985 г.). - Киев: Ин-т пробл. моделирования в энергетике АН УССР, 1985. - Ч. 2. - С. 44-45.
2. Платонов Б. А., Ромейко В. А., Сигорский П. В. Обработка данных в сети океанографических центров // Тез. докл. Одиннадцатого Всесоюз. семинара по вычислительным сетям (Рига, 1986 г.). - 1986. - Ч. 3. - С. 338-343.
3. Ромейко В. А., Сигорский П. В. Представление информации в системах обработки изображений // Автоматизированные системы обработки изображений: Тез. докл. II Всесоюз. конф. (Львов, 1986 г.). - М.: Наука, 1986. - С. 219-221.
4. Задорожный В. В., Сигорский П. В. Параллельные процедуры для задач классификации // Распараллеливание обработки информации: Тез.

докл. VI Всесоюз. школы-семинара (Львов, 1987 г.). — Львов : Физико-мех. ин-т им. Г. В. Карпенко АН УССР, 1987. — Ч. 1. — С. 140—141.

5. Сигорский П. В. Методика построения и расчета интегрированных систем обработки информации // Там же. — С. 189—190.

6. Сигорский П. В. Статистические характеристики интегрированной системы обработки океанографических данных // Вычисл. и прикл. математика. — 1987. — Вып. 62. — С. 113—122.

7. Сигорский П. В. Аналитическое моделирование систем интеграции пакетов прикладных программ // Вопр. кибернетики. — 1987. — Вып. 123. — С. 151—161.

Подп. в печ. 21.08.87. БФ 11462. Формат 60×84/16. Бум. писч. Офс. печ. Усл. печ. л. 0,93. Усл. кр.-отт. 0,93. Уч.-изд. л. 1,0. Тираж 100 экз. Заказ 1705. Бесплатно.

РИО Института кибернетики имени В. М. Глушкова АН УССР
252207 Киев 207, проспект Академика Глушкова, 20