

1553029

На правах рукописи

СОРОКИН СЕРГЕЙ ВИКТОРОВИЧ

**ПАРАМЕТРИЧЕСКИЕ АЛГОРИТМЫ РЕШЕНИЯ ЗАДАЧ
РАСПРЕДЕЛЕНИЯ ЦЕЛОЧИСЛЕННЫХ РЕСУРСОВ БОЛЬШОЙ
РАЗМЕРНОСТИ В СИСТЕМАХ РЕАЛЬНОГО ВРЕМЕНИ**

Специальность 05.13.18 - Теоретические основы математического
моделирования, численные методы и
комплексы программ

**Автореферат
диссертации на соискание ученой степени кандидата
технических наук**



Москва - 1996

Работа выполнена в научно-техническом центре "Модуль"

Научный руководитель - доктор технических наук,
академик РА РАН М.Ф. Яфраков

Официальные оппоненты: доктор технических наук,
профессор Ц.Г. Литовченко,
кандидат технических наук
В.С. Козлов

Ведущая организация - Научно-исследовательский
институт радиоприборостроения

Защита состоится " 11 " *февраль* 1997 г. в *11 часов*
на заседании диссертационного совета Д 053.15.12 при Московском
государственном техническом университете им. Н.Э. Баумана по
адресу: 107005, Москва, 2-я

С диссертацией мо-
имени Н.Э. Баумана

Автореферат разосла:

Ученый секретарь
диссертационного со-
д. ф.-м. н., профессор

1553029



1553029

Сорокин С.В. Параметрическ

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность работы

Диссертационная работа посвящена разработке и исследованию приближенных параметрических алгоритмов решения задач распределения целочисленных ресурсов (РЦР). В таких задачах требуется распределить ограниченное количество неделимых ресурсов по нескольким подсистемам (объектам назначения) с целью достижения максимальной эффективности системы в целом.

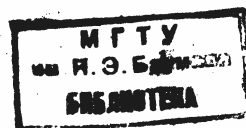
Наиболее часто задачи РЦР возникают при управлении различными техническими системами (вычислительными комплексами, информационными сетями, системами обороны и обеспечения надежности). Современный уровень развития техники и потребности практики накладывают жесткие временные ограничения на управление такими системами. При этом соответствующие задачи РЦР часто имеют большую размерность (десятки-сотни ресурсов и объектов назначения) и должны решаться с высокой точностью в реальном масштабе времени (за доли секунд-секунды).

Существующие алгоритмы решения задач РЦР не удовлетворяют в полной мере современным требованиям по точности и трудоемкости и в ряде случаев не позволяют получать оптимальные или близкие к оптимальным решения задач большой размерности за требуемое (желаемое) время. В связи с этим актуальной является проблема разработки новых алгоритмов, ориентированных на современную и перспективную вычислительную технику, легко адаптируемых к различным условиям применения и позволяющих получать высокоточные решения задач РЦР с меньшими, чем у существующих аналогов временными затратами.

Предпосылками к созданию таких алгоритмов являются: результаты современных исследований в области комбинаторной оптимизации (работы Силина В.Б., Спесивцева А.В., Емеца О.А., Хохлюка В.И., Генса Г.В., Левнера Е.В. и других авторов), интенсивное развитие теории параллельных вычислений и искусственных нейронных сетей, широкое распространение компьютеров с параллельной архитектурой.

Целью работы является

- теоретическое и экспериментальное обоснование построения параметрических алгоритмов решения задач РЦР с варьируемыми характеристиками точности и трудоемкости;



- разработка приближенных параметрических алгоритмов, ориентированных на современную и перспективную вычислительную технику и позволяющих сократить временные затраты на решение задач РЦР большой размерности при сохранении высокой точности получаемых решений.

Методы исследования

В ходе теоретических исследований, проведенных в работе, использовались понятия и методы исследования операций и математического программирования, комбинаторного анализа, теории алгоритмов, теории параллельных вычислений и искусственных нейронных сетей.

Научная новизна работы состоит в следующем:

- проведено исследование комбинаторных свойств задач РЦР и созданы основы синтеза параметрических алгоритмов их решения с варьируемыми характеристиками точности и трудоемкости;
- разработан универсальный комбинаторный метод решения задач РЦР, позволяющий в зависимости от предъявляемых требований синтезировать алгоритмы различной точности и трудоемкости, включая алгоритмы с гарантированными оценками погрешности;
- разработано два параметрических семейства алгоритмов, позволяющих получать высокоточные решения задач РЦР при меньших временных затратах, чем известные аналоги; проведено исследование алгоритмов, получены оценки их точности и трудоемкости;
- предложен способ распараллеливания разработанных алгоритмов и модель нейросетевой вычислительной системы, на которой данный способ может быть эффективно реализован;
- предложен вариант параллельной реализации разностных алгоритмов на нейрокомпьютерной рабочей станции (НКРС), содержащей плату-нейровычислитель с четырьмя сигнальными процессорами;
- при реализации алгоритмов на НКРС установлен близкий к линейному рост ускорения вычислений при увеличении числа используемых процессоров.

Практическая ценность и реализация результатов работы

Полученные и представленные в диссертационной работе результаты позволяют с высокой точностью решать задачи РЦР большой размерности в условиях жестких временных ограничений (за доли секунды) и могут использоваться при управлении техническими системами, работающими в реальном или близком к реальному масштабе времени. В частности, предложенные алгоритмы могут использоваться для решения задач распределения

вычислительных ресурсов, задач целераспределения, возникающих в процессе управления системами обороны.

Описанные в работе алгоритмы были включены в состав программно-алгоритмического обеспечения, разрабатываемого в целях совершенствования вычислительно-моделирующей базы систем военно-технического назначения, и применялись в процессе математического моделирования различных вариантов этих систем.

Результаты работы внедрены в НИИРП и АО НТЦ "Модуль", о чем свидетельствуют прилагаемые акты.

Апробация работы

Основные положения диссертационной работы были доложены на международной конференции "Эволюция инфосферы" (1995г.) и постоянно действующем научно-техническом семинаре в НТЦ "Модуль".

По теме диссертации опубликовано 6 статей. Материалы работы отражены в шести научно-технических отчетах.

Структура и объем работы

Диссертация состоит из введения, четырех глав, заключения, библиографии, включающей 77 наименований, и трех приложений. Общий объем - 137 страниц.

На защиту выносятся:

- результаты теоретического и экспериментального исследования целевых функций задач РЦР, заданных на специально введенных метрических комбинаторных пространствах;
- универсальный комбинаторный метод -решения задач РЦР, являющийся основой для построения параметрических алгоритмов различной точности и трудоемкости;
- параметрические алгоритмы решения задач РЦР и результаты их экспериментального исследования;
- способ распараллеливания разработанных алгоритмов и его нейросетевая реализация.

СОДЕРЖАНИЕ РАБОТЫ

Во введении дается общая характеристика диссертационной работы, кратко описывается ее содержание и основные полученные результаты.

В первой главе приводятся постановки общей и частных задач РЦР, проводится анализ и классификация методов и алгоритмов их

решения, определяются основные цели работы и пути их достижения.

В общем виде задача РЦР формулируется следующим образом. Имеется m объектов назначения и n ресурсов. Числовые характеристики ресурсов задаются матрицей $P = \|p_{ij}\|$ размера $n \times m$, где p_{ij} - эффективность использования ресурса i на объекте j . Распределение ресурсов по объектам определяется бинарной матрицей $X = \|x_{ij}\|$ размера $n \times m$, называемой планом распределения (назначения). При этом $x_{ij}=1$, если i -й ресурс назначен на j -й объект и $x_{ij}=0$ в противном случае. Каждый ресурс может быть назначен только на один объект, т.е. $\sum_{j=1}^m x_{ij} \leq 1, \forall i$.

Требуется найти такую матрицу X^* , что $F(P, X^*) = \max_{X \in \Omega} F(P, X)$, где: Ω - множество допустимых планов (решений) задачи, а F - функционал (целевая функция), представимый в виде

$$F(P, X) = \sum_{j=1}^m f(p_{1j}, p_{2j}, \dots, p_{nj}, x_{1j}, x_{2j}, \dots, x_{nj}).$$

Комбинаторная специфика задач РЦР часто не позволяет находить их точное решение за отведенное время. В этом случае ищется приближенное решение, т.е. такая матрица $\tilde{X} \in \Omega$, что

$$\left| \max_{X \in \Omega} F(P, X) - F(P, \tilde{X}) \right| \leq \delta, \delta > 0.$$

Если числовые характеристики ресурсов одинаковы, то такие ресурсы принято называть однородными (неразличимыми) в отличие от неоднородных ресурсов, которые имеют различные числовые характеристики. Наиболее сложными в вычислительном плане являются задачи распределения неоднородных ресурсов, многие из которых принадлежат к классу NP-трудных проблем.

Примерами задач РЦР являются задача целераспределения для неоднородных средств поражения (WAP) и задача о назначениях (AP).

Задача WAP состоит в распределении n разнотипных средств поражения по m целям противника по критерию минимума ожидаемого ущерба (или по критерию максимума предотвращенного ущерба). При этом известно, что j -я цель в случае ее пропуска (непоражения) наносит ущерб в β_j единиц ($j = \overline{1, m}$), а вероятность поражения j -й цели i -м средством ($i = \overline{1, n}$) равна $p_{ij} = 1 - q_{ij}$. Считается, что средства поражения действуют независимо друг от друга, причем каждое средство может поразить только одну цель. Формально задача WAP записывается в виде:

$$\sum_{j=1}^m \beta_j (1 - \prod_{i=1}^n q_{ij}^{x_{ij}}) \rightarrow \max,$$

$$\sum_{j=1}^m x_{ij} = 1, i = \overline{1, n},$$

$$\beta_j \geq 0, \quad 0 < q_{ij} \leq 1, \quad x_{ij} \in \{0, 1\},$$

$$i = \overline{1, n}, \quad j = \overline{1, m}.$$

Важным частным случаем задачи WAP является задача AP, наиболее известная из литературы содержательная постановка которой состоит в следующем. Имеется n исполнителей, которые могут выполнять n работ. Известна производительность c_{ij} выполнения i -м работником j -й работы ($i, j = \overline{1, n}$). Необходимо так назначить исполнителей на работы, чтобы все работы выполнялись, каждый работник выполнял только одну работу, а суммарная производительность была максимальной.

Формально задача записывается следующим образом:

$$\sum_{i=1}^n \sum_{j=1}^n c_{ij} x_{ij} \rightarrow \max,$$

$$\sum_{i=1}^n x_{ij} = 1, j = \overline{1, n},$$

$$\sum_{j=1}^n x_{ij} = 1, i = \overline{1, n},$$

$$x_{ij} \in \{0, 1\}, \quad i = \overline{1, n}, j = \overline{1, n}$$

Задачи AP и WAP имеют большое теоретическое и прикладное значение, поэтому основные результаты диссертационной работы демонстрируются, как правило, на примере этих задач.

Алгоритмы решения задач РЦР в диссертационной работе классифицируются по двум основным критериям: точность и трудоемкость. Трудоемкость алгоритмов оценивается двумя способами - теоретическим и эмпирическим (экспериментальным). Первый способ основан на общей теории сложности вычислений, второй (эмпирический) - на замерах реального времени решения задачи при реализации на конкретной ЭВМ.

Проведенные в диссертационной работе исследования, а также многочисленные факты, известные из литературы (работы Спесивцева А.В., Генса Г.В., Хохлюка В.И., Гэрри М., Джонсона Д. и других) свидетельствуют о том, что точные алгоритмы (основанные на методе ветвей и границ, принципе оптимальности Беллмана, методах отсечения и другие) имеют высокую трудоемкость и в большинстве случаев не могут использоваться при решении задач

РЦР большой размерности в условиях жестких временных ограничений. В связи с этим целесообразным является применение приближенных алгоритмов, позволяющих за допустимое время находить решения, близкие к оптимальным.

Анализ известных приближенных алгоритмов показывает, что большинство из них имеют существенные недостатки. Некоторые из этих недостатков носят частный характер, другие являются общими и в той или иной степени присущи большинству известных алгоритмов. Частными недостатками являются высокая погрешность и трудоемкость, узкая специализированность, необходимость округления полученных нецелочисленных решений. Часто указанные недостатки являются следствием отказа от условий целочисленности и использования традиционных методов математического программирования (симплекс-метода, метода множителей Лагранжа, методов спуска). К общим недостаткам следует отнести изначальную ориентацию на ЭВМ с последовательной архитектурой (что не позволяет в полной мере использовать преимущества современных многопроцессорных ЭВМ), ограниченные возможности варьирования точности и трудоемкости.

Указанные недостатки затрудняют практическое использование известных алгоритмов при решении задач большой размерности в условиях дефицита времени и в ряде случаев не позволяют добиться требуемых (либо желаемых) точности и трудоемкости. В связи с этим необходимо разрабатывать новые алгоритмы, адаптируемые к различным условиям применения и позволяющие значительно увеличить размерности задач РЦР, решаемых с высокой точностью за допустимое время. Разработка таких алгоритмов, а также их теоретическое и экспериментальное обоснование является основной целью данной работы.

Для достижения поставленной цели в диссертационной работе решаются следующие задачи:

- исследование комбинаторных свойств задач РЦР и создание основ синтеза параметрических алгоритмов их решения с варьируемыми характеристиками точности и трудоемкости;
- разработка алгоритмов и их реализация на различных видах вычислительной техники;
- теоретическое и экспериментальное исследование алгоритмов, сравнение с известными аналогами;
- разработка моделей вычислительных систем, на которых созданные алгоритмы могут быть эффективно реализованы.

Первая глава диссертации заканчивается кратким описанием подходов к решению перечисленных задач.

Вторая глава диссертационной работы посвящена исследованию специфических комбинаторных свойств задач РЦР и разработке универсального комбинаторного метода их решения.

Исследование основных свойств задач РЦР проводится в специально введенных метрических комбинаторных пространствах. Элементами (точками) этих пространств являются n -мерные вектора вида $x = (x_1, x_2, \dots, x_n)$, где номер элемента соответствует номеру распределяемого ресурса, а его значение - номеру объекта, на который этот ресурс назначается. Расстояние между точками x и y комбинаторного пространства определяется функцией

$$\rho(x, y) = \sum_{i=1}^n \text{sign}(|x_i - y_i|), \text{ где } \text{sign}(z) = \begin{cases} 0, & z = 0, \\ 1, & z \neq 0. \end{cases}$$

Совокупность всех точек y комбинаторного пространства, таких, что $\rho(x, y) \leq \varepsilon$ называется ε -окрестностью (или окрестностью радиуса ε) точки x и обозначается $O(x; \varepsilon)$. При этом $\varepsilon \in \{0, 1, 2, \dots, n\}$. Размером окрестности называется число входящих в нее точек. Невырожденными считаются окрестности размером больше единицы.

Точку $\bar{x}$ комбинаторного пространства назовем точкой локального экстремума целевой функции F , если она удовлетворяет условию: $F(\bar{x}) = \text{extr} F(x)$, где окрестность $O(\bar{x}; r)$ является невырожденной.

Введенные понятия позволяют формулировать задачи РЦР как задачи оптимизации на комбинаторных пространствах и использовать свойства этих пространств и заданных на них целевых функций при разработке алгоритмов.

В процессе исследования комбинаторных свойств задач РЦР получены формулы и алгоритмы для пересчета и перечисления элементов комбинаторных пространств и локальных окрестностей фиксированного радиуса, оценки мощности множества различных значений целевых функций задач РЦР, оценки относительной погрешности для точек локального экстремума и алгоритмов ограниченного перебора в комбинаторных пространствах. В частности, для задачи АР доказано, что каждой точке локального экстремума соответствует приближенное решение с

$$\text{гарантированной оценкой погрешности } q = \frac{F(x^*) - F(x)}{F(x^*)} 100\% \leq 50\%$$

(здесь $F(x^*)$ и $F(x)$ соответственно оптимальное значение целевой функции и значение, достигаемое в точке локального экстремума). Приводится пример, показывающий, что данная оценка является достижимой.

Экспериментальное исследование локальных экстремумов задач AP и WAP состояло в определении их количества, размеров областей сходимости, экстремальных значений целевых функций.

В результате проведенного исследования было установлено следующее.

1. Целевые функции рассматриваемых задач имеют многоэкстремальный характер.

2. Количество локальных экстремумов и размеры соответствующих областей сходимости существенно зависят от размерности задачи и исходных данных. Как правило, количество локальных экстремумов мало по сравнению с мощностью комбинаторного пространства.

3. Наибольшая область сходимости соответствует глобальному экстремуму.

4. Области сходимости имеют тенденцию к увеличению при приближении локальных экстремумов к глобальным.

Исследования проводились для задач с равномерно распределенными и нормально распределенными исходными данными.

Полученные во второй главе диссертационной работы результаты теоретического и экспериментального исследования легли в основу разработанного комбинаторного метода решения задач РЦР.

Метод состоит в последовательном выполнении следующих шагов.

1. Определить для решаемой задачи РЦР комбинаторное пространство Ω .

2. Сформулировать решаемую задачу как задачу оптимизации на комбинаторном пространстве:

найти (а) $x^* = \arg \min_{x \in \Omega} F(x)$, (б) $\bar{x} \in \Omega$, $\left| \min_{x \in \Omega} F(x) - F(\bar{x}) \right| \leq \delta$, $\delta > 0$.

3. Определить одну или несколько исходных точек x^j комбинаторного пространства (опорных планов).

4. Задав радиус $r_{\min}$ для каждого из опорных планов найти точку локального экстремума

$$\bar{x}^j = \arg \min_{x \in O(\bar{x}^j, r_{\min})} F(x).$$

5. В зависимости от требований, предъявляемых к точности и трудоемкости решения задачи, либо закончить вычисления, приняв за решение лучший из найденных локальных экстремумов, либо увеличить радиус рассматриваемых окрестностей и продолжить поиск.

В случае необходимости поиск локальных экстремумов может быть прерван на любом промежуточном этапе.

Описанный метод является основой для построения параметрического семейства алгоритмов решения задач РЦР, где варьируемыми параметрами являются радиусы просматриваемых окрестностей и количество опорных планов. При этом имеется возможность построения алгоритмов с гарантированными оценками погрешности.

В случае использования нескольких опорных планов алгоритмы локального поиска для каждого из них могут выполняться параллельно, что позволяет значительно сократить время решения задачи.

Проведенное экспериментальное исследование показало, что точность и трудоемкость алгоритмов разработанного метода за счет параметризации может варьироваться в широком диапазоне. При этом даже простейшие из таких алгоритмов в большинстве случаев позволяют получать решения, близкие к оптимальным. В частности, было установлено, что для задачи АР размерности $n \leq 20$ в случае однократного использования алгоритма наискорейшего спуска с радиусом просматриваемой на каждом шаге окрестности $r=2$ среднестатистическая относительная погрешность получаемых решений не превосходит 2,01%.

В третьей главе диссертационной работы рассматриваются так называемые разностные алгоритмы, которые могут иметь как самостоятельное значение, так и использоваться в сочетании с алгоритмами локального поиска в комбинаторных пространствах (на этапе получения опорных планов).

Основные результаты, полученные в данной главе демонстрируются на примере задачи WAP и ее различных частных случаев.

В общем виде разностные алгоритмы решения задачи WAP могут быть записаны следующим образом.

1. Принять $x_{ij} := 0$, $i = \overline{1, n}$, $j = \overline{1, m}$, $s = 0$.

2. Найти значения индексов k и r в соответствии с критерием:

$$\Delta_{kr} = \max \Delta_{ij}, \text{ где } \Delta_{ij} = \beta_j p_{ij} + L_s(X, i, j), \quad i = \overline{1, n}, j = \overline{1, m}$$

(L_s - некоторая функция, где $X = \|x_{ij}\|$ - переменные, i, j - параметры, s - номер шага).

3. Пересчитать значения β_r, x_{kr}, s :

$$\beta_r := \beta_r q_{kr}, \quad x_{kr} := 1, \quad s = s + 1.$$

4. Положить $p_{kj} = 0$, $j = \overline{1, m}$.

5. Если $s=n$, то закончить вычисления; если нет, перейти к пункту 2.

Частные алгоритмы получаются из общей алгоритмической схемы путем выбора конкретной функции L_s .

Первоначально в работе анализируются два известных из литературы разностных алгоритма: быстродействующий "жадный" алгоритм, получаемый из общей алгоритмической схемы при $L_s = 0$, и алгоритм Берзина, для которого

$$L_s = \frac{\beta_j p_{ij}}{q_{ij}} \prod_{r=1}^n q_{rj} - \sum_{i=1}^m \frac{\beta_i p_{ii}}{q_{ii}} \prod_{r=1}^n q_{ri}.$$

Показаны преимущества и недостатки каждого из алгоритмов.

Основные недостатки алгоритмов состоят в следующем: при решении задач распределения неоднородных ресурсов "жадный" алгоритм часто приводит к большой относительной погрешности (более 10%), а алгоритм Берзина имеет высокую трудоемкость, связанную с необходимостью многократного вычисления значений функции L_s .

Далее рассматривается разработанное автором параметрическое семейство разностных алгоритмов, обобщающее как ранее известные, так и новые алгоритмы. В алгоритмах этого семейства величины Δ_{ij} вычисляются по формуле

$$\Delta_{ij} = \beta_j p_{ij} - \alpha_s \sum_{k \neq j} \beta_k p_{ik},$$

где α_s - параметры, выбираемые в зависимости от вида и размерности решаемой задачи.

Предложен способ задания параметров α_s , позволяющий учитывать требования, предъявляемые к точности и трудоемкости алгоритма. В простейшем случае при решении задачи WAP можно использовать постоянные для каждого шага параметры α_s ,

вычисляемые по формуле $\alpha_s = \frac{1}{m-1}$. Теоретически трудоемкость

алгоритмов оценивается величиной $O(n^2 m)$, что совпадает с аналогичной оценкой для "жадного" алгоритма. При корректном задании параметров α_s разработанные алгоритмы превосходят по точности значительно более трудоемкий алгоритм Берзина (относительная погрешность получаемых решений редко превосходит 3%).

Разработанные алгоритмы устраняют недостатки известных аналогов и могут использоваться для решения задач большой размерности в условиях дефицита времени.

В случае необходимости решения, получаемые с помощью разностных алгоритмов, могут быть улучшены алгоритмами локального поиска в комбинаторных пространствах. Даже простейшие алгоритмы локального поиска позволяют в 1,5-2 раза снизить среднестатистическую погрешность получаемых решений при относительно малом увеличении суммарных временных затрат.

На рис. 1 представлены экспериментально полученные графики зависимости среднестатистической погрешности q (%) от размерности задачи АР для одного из алгоритмов разработанного параметрического семейства без локального поиска (график 1) и с локальным поиском в окрестностях радиуса $r=2$ (график 2). При использовании алгоритма локального поиска суммарные временные затраты увеличивались не более чем на 10-15%.

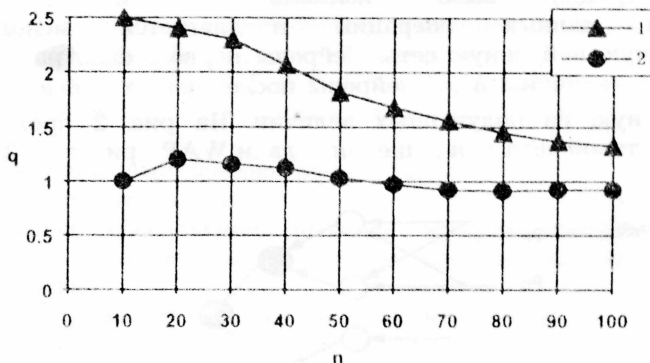


Рис. 1

Четвертая глава диссертационной работы посвящена параллельной реализации разностных алгоритмов на компьютерах с нейросетевой архитектурой.

Под искусственной нейронной сетью (ИНС) понимается параллельная вычислительная система, состоящая из совокупности взаимосвязанных процессорных элементов (нейронов), каждый из которых выполняет две простейшие операции - взвешенное суммирование и линейное либо нелинейное преобразование.

Как правило, нейрон имеет один выход и несколько входов. На входы нейрона подаются сигналы с выходов других, связанных с данным нейроном, а также внешние воздействия. Сигнал на выходе

нейрона определяется выражением $y = F(\sum_{i=1}^n t_i x_i)$, где $(x_1, x_2, \dots, x_n)$ - вектор входных сигналов, $(t_1, t_2, \dots, t_n)$ - вектор весовых коэффициентов (синаптических связей), F - функция активации нейрона. В качестве функций активации обычно используются сигмоидные, линейные кусочно-линейные функции.

В настоящее время ИНС аппаратно реализуются на электронных и оптических элементах или эмулируются с использованием специализированных многопроцессорных плат.

При реализации разностных алгоритмов на ЭВМ с последовательной архитектурой основные вычислительные затраты приходится на многократное выполнение двух операций - вычисление величин Δ_{ij} (взвешенное суммирование элементов строки матрицы) и выбор максимальной из этих величин. Для ускорения данных операций предлагается использовать многослойную нейронную сеть. Нейроны первого слоя параллельно вычисляют величины Δ_{ij} , а нейроны последующих слоев выбирают максимальную из полученных величин. На рис. 2 представлена структура такой сети для решения задачи WAP при $n=m=2$.

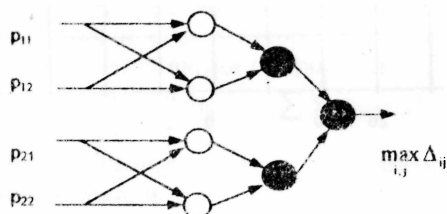


Рис. 2

Операция выбора максимального элемента выполняется с помощью элементарных нейронных ансамблей, реализующих следующую простую формулу для выбора максимального из двух чисел: $\max(a, b) = \frac{a + b}{2} + \frac{|a - b|}{2}$. На рисунке эти ансамбли выделены темным цветом.

Использование вычислительных систем с подобной нейросетевой архитектурой позволяет в $O(nm/\ln(mn))$ раз ускорить вычисления за счет распараллеливания.

Проведенный в работе сравнительный анализ показывает, что при решении задач РЦР предложенная нейросетевая модель организации параллельных вычислений предпочтительнее известных моделей Хопфилда, так как является более простой и гибкой, требует меньших аппаратных затрат, ориентирована на алгоритмы с лучшими точностными характеристиками.

Наибольшее быстродействие разностных алгоритмов достигается в случае полностью параллельной аппаратной реализации соответствующих ИНС. Однако существующие на

сегодняшний день технологические ограничения не позволяют полностью аппаратно реализовывать ИНС для решения задач большой размерности. В связи с этим одним из наиболее перспективных и интенсивно развиваемых в последние годы направлений нейрокомпьютинга является программно-аппаратная эмуляция ИНС с применением многопроцессорных плат-ускорителей нейровычислений.

В диссертационной работе предложен и осуществлен на практике вариант параллельной реализации разработанных разностных алгоритмов на нейрокомпьютерной рабочей станции (НКРС), содержащей плату-нейровычислитель с четырьмя сигнальными процессорами TMS320C40. В состав НКРС входит:

- управляющая ЭВМ (персональный компьютер IBM PC 486 DX4-100);
- специализированный нейровычислитель;
- средства комплексирования;
- программное обеспечение (штатное программное обеспечение персонального компьютера + базовое программное обеспечение нейровычислителя).

Общая производительность нейровычислителя достигает 1100 MOPS и 200 MFLOPS. Упрощенная схема структурной организации НКРС представлена на рис. 3.



Рис. 3

При программной реализации алгоритмов на НКРС использовался принцип геометрического параллелизма, согласно которому исходная матрица построчно разбивалась на четыре части, каждая из которых обрабатывалась отдельным процессором. Применительно к нейронным сетям геометрический параллелизм означает эмулирование каждым из процессоров группы нейронов.

В результате проведенного экспериментального исследования был зафиксирован близкий к линейному рост ускорения вычислений при увеличении числа параллельно работающих

процессоров, а также 5-7 кратное снижение временных затрат при реализации алгоритмов на плате-ускорителе по сравнению с последовательной реализацией на компьютере IBM PC 486 DX4-100.

На рис. 4 представлены экспериментально полученные графики трудоемкости t (с) одного из алгоритмов решения задачи WAP при $n=m$ для различных вариантов реализации.

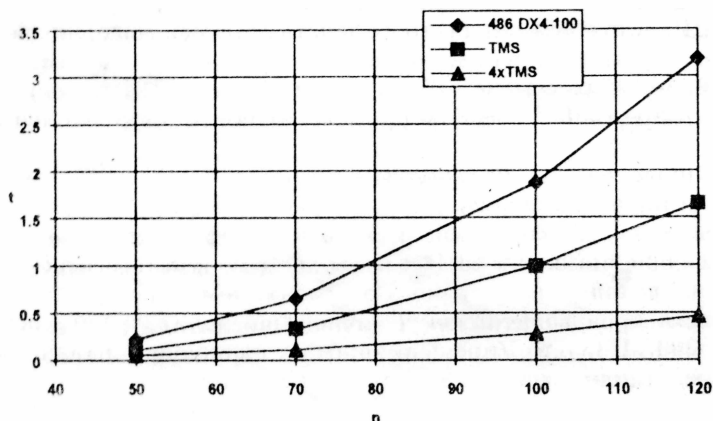


Рис. 4

Полученные данные свидетельствуют о возможности значительного ускорения вычислений за счет распараллеливания разработанных алгоритмов и целесообразности их использования при решении задач РЦР большой размерности в условиях дефицита времени.

В заключении сформулированы основные результаты, полученные в работе.

ОСНОВНЫЕ РЕЗУЛЬТАТЫ

1. Проведено теоретическое и экспериментальное исследование целевых функций задач РЦР и их экстремумов на специально введенных метрических комбинаторных пространствах. Получены формулы и алгоритмы для пересчета и перечисления элементов комбинаторных пространств и локальных окрестностей фиксированного радиуса, оценки погрешности в точках локального экстремума и мощности множества этих точек. Для задач о назначении и целераспределения установлена зависимость между

значениями целевых функций в точках локального экстремума и размером соответствующих им областей сходимости.

2. На основании проведенных исследований предложен универсальный метод решения задач РЦР, позволяющий синтезировать алгоритмы с варьируемыми характеристиками точности и трудоемкости, включая алгоритмы с гарантированными оценками погрешности.

3. Разработано и реализовано в виде пакета прикладных программ два параметрических семейства алгоритмов решения задач РЦР. Алгоритмы первого семейства (разностные) позволяют за малое время находить решения, близкие к оптимальным. Алгоритмы второго семейства позволяют улучшать эти решения и могут быть прерваны на любом промежуточном этапе в зависимости от накладываемых временных ограничений.

4. Проведено исследование алгоритмов, выработаны рекомендации по их практическому использованию. Показано, что разработанные алгоритмы могут быть адаптированы к различным условиям решения задач РЦР и позволяют получать высокоточные решения при меньших временных затратах, чем известные аналоги.

5. Разработана модель нейросетевой вычислительной системы, позволяющей значительно ускорить решение задач РЦР созданными алгоритмами за счет распараллеливания.

6. Предложен и осуществлен на практике вариант параллельной реализации разностных алгоритмов на НКРС, содержащей плату-ускоритель с 4-мя процессорами. При решении на НКРС задач о назначениях и целераспределения установлено близкое к линейному снижение временных затрат с увеличением числа используемых процессоров.

7. Экспериментальные данные, полученные на НКРС, показывают, что разработанные алгоритмы позволяют находить близкие к оптимальным решения задач РЦР большой размерности ($n, m > 100$) за доли секунды.

Полученные результаты позволяют значительно сократить время решения задач РЦР большой размерности при сохранении высокой точности получаемых решений и могут использоваться в технических системах, работающих в реальном времени.

ПУБЛИКАЦИИ ПО ТЕМЕ ДИССЕРТАЦИИ

1. Сорокин С.В. Комбинаторный подход к решению задачи рационального распределения ограниченных ресурсов // Отраслевой научно-технический журнал. Сер. СП. - 1990. - Вып. 1(5). - С. 69-75.

2. Криюлев С.Н., Сорокин С.В. Разработка и анализ алгоритмов оптимизации целочисленного распределения ресурсов в моделях сложных систем // Отраслевой научно-технический журнал. Сер. СОИУ. - 1991. - Вып. 3. -С. 55-72.
3. Сорокин С.В. Решение задачи о назначениях с использованием нейроподобных сетей Хопфилда // Известия вузов. Приборостроение. - 1994. - Т. 37, N3-4. -С. 57-59.
4. Сорокин С.В. Разностные алгоритмы распределения неоднородных целочисленных ресурсов // Теория и системы управления. - 1995. - N6. -С. 218-222.
5. Сорокин С.В. Нейросетевая реализация разностных алгоритмов распределения целочисленных ресурсов // Известия вузов. Приборостроение. - 1995. - Т. 38, N1-2. -С. 28-31.
6. Сорокин С.В. Решение задач распределения целочисленных ресурсов на нейрокомпьютерной рабочей станции //Известия вузов. Приборостроение. -1996. -Т. 39, N1. -С. 69-70.