УДК 001.89:518.5

ЭВРИСТИЧЕСКИЙ ПОДХОД К РЕШЕНИЮ ДИНАМИЧЕСКИХ ЗАДАЧ РАСПРЕДЕЛЕНИЯ РЕСУРСОВ

В. Н. БУРКОВ, С. Е. ЛОВЕЦКИЙ

(Москва)

Рассматриваются динамические задачи распределения ресурсов по множеству работ (операций) в системах сетевого планирования и управления. При этом предполагается, что ресурсы после использования на одной работе можно использовать для выполнения других работ данного класса (например, машины, оборудование, люди и т. д.). Предлагается эвристический подход к решению динамических задач. В ряде случаев алгоритм дает оптимальное решение. Получены оценки степени оптимальности.

Задачи, связанные с распределением ограниченных ресурсов по множеству операций, представляют в настоящее время большой интерес. В работе рассматриваются динамические задачи распределения ресурсов по множеству работ (операций) в системах сетевого планирования и управления (ССПУ). Под динамическими задачами распределения ресурсов будем понимать задачи, в которых ресурсы после использования на одной операции можно использовать для выполнения других операций данного класса (например, машины, оборудование, люди и т. д.). В процессе выполнения операций ресурсы переходят с операции на операцию, образуя поток по множеству операций. Существуют различные методы решения динамических задач (методы теории графов, линейного программирования, динамического программирования и т. д.). Однако все эти методы в большинстве случаев неприменимы к задачам большой размерности, так как объем вычисления быстро возрастает с увеличением размерности задачи («проклятие размерности» по выражению Беллмана). Поэтому все большее внимание уделяется созданию эвристических алгоритмов, позволяющих получать численные решения сложных задач.

Ниже предлагается эвристический подход к решению динамической задачи распределения ресурсов. Дается постановка задачи и краткий обзор методов ее решения, получены оценки для времени выполнения комплекса операций, приводится формальное описание алгоритма. Наконец, рассматриваются приемы, позволяющие в некоторых случаях улучшить решение, полученное в результате применения алгоритма.

1. Постановка задач

Задача I. Требуется выполнить комплекс, состоящий из n операций. Будем обозначать операции вершинами графа, причем от вершины i идет дуга к вершине j, если операция A_j не может быть начата, пока не закончена операция A_i (говорят, что задана сопряженная сеть).

Множество операций разбито на k классов (в данной работе рассматривается случай $k = 1$). Для каждого класса задано количество ресурсов N_j $(j = 1, 2, \ldots, k)$, предназначенных для выполнения операций данного класса. Обозначим через W_i объем i-й операции (в единицах «ресурсы $\times$ $\times$ время»). Будем считать, что скорость выполнения операции (объем операции, выполненный в единицу времени) является линейной функцией количества ресурсов на ней в некоторых пределах, т. е.

$$w_i(t) = a_i + b_i u_i(t), \qquad \alpha_i \leqslant u_i(t) \leqslant \beta_i \qquad (i = 1, 2, \ldots, n),$$

где $w_i(t)$ — скорость выполнения i-й операции в момент t, $u_i(t)$ — количество ресурсов на i-й операции в момент t.

В дальнейшем рассматривается случай $a_i = 0$ (не уменьшая общности, можно считать $b_i = 1$). Требуется распределить ресурсы по операциям так, чтобы время выполнения комплекса было минимальным.

Задача II. Требуется выполнить комплекс из n операций (очередность их выполнения задана в виде сетевого графика). Все операции разбиты на k классов. Для каждого класса задано количество ресурсов N_j $(j = 1, 2, \ldots, k)$, предназначенное для выполнения операций данного класса.

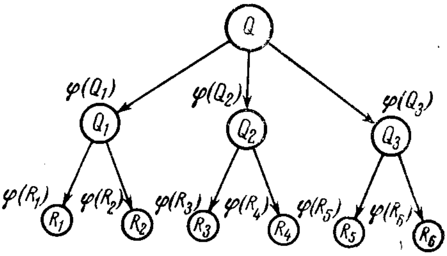


Рис. 1

Для выполнения i-й операции требуется β_i ед. ресурсов, которые выполняют ее за время τ_i, причем прерывать ее выполнение запрещается. Требуется распределить ресурсы по работам так, чтобы выполнить комплекс за минимальное время.

Легко убедиться, что задача II получается из задачи I при $\alpha_i = \beta_i$ $(i = 1, 2, \ldots, n)$ (при этом время выполнения $\tau_i = W_i / \beta_i$), если запретить прерывать начатую работу. Задача II является типичной задачей теории расписаний.

В [1] предложен алгоритм ее решения, аналогичный рассматриваемому в данной работе. Правило распределения следующее: в первую очередь выполняются операции с меньшим резервом времени. Частный случай задачи II (все времена равны и все операции одного класса) рассмотрен в [2], где она называется задачей сбалансирования сборочной линии (line-balance task).

В [2] предложен алгоритм решения задачи II и показано, что для графа вида дерева алгоритм дает оптимальное решение. К частному случаю задачи II следует отнести и известную задачу обработки деталей на станках. В последнее время предложены довольно эффективные методы ее приближенного решения [3, 4]. Анализ этих методов, а также алгоритмов [1, 2] и др. показывает, что они укладываются в довольно общую схему решения комбинаторных задач — метод ветвлений (в [5] этот метод назван методом функций предпочтения). Кратко остановимся на основных идеях этого метода.

Обозначим через Q множество решений комбинаторной задачи. Пусть задана некоторая функция оценки $\varphi(Q_i)$ для подмножеств $Q_i \subset Q$. Определим процесс ветвления как процесс разбиения множества решений Q на подмножества Q_i, каждого Q_i на подмножества R_i и т. д. и будем выбирать каждый раз подмножество с минимальным (максимальным) значением функции оценки. Процессу ветвления можно поставить в соответствие дерево ветвления (рис. 1), так что выбор некоторого подмножества соответствует выбору направления движения по дереву. Отметим, что

$\varphi(Q_i)$ может соответствовать вероятности выбора данного подмножества (метод рандомизации [5]). Иногда в качестве функции оценки берется нижняя граница значений минимизируемого функционала для данного подмножества. Тогда, очевидно, можно не рассматривать далее подмножества, для которых значение нижней границы превосходит значения функционала для некоторого (уже полученного) решения (метод ветвей и границ [6, 3]).

Вернемся к задачам распределения ресурсов. Пусть $F(t)$ — множество операций, которые можно выполнять (или выполняются) в момент t. Это множество называется фронтом операций ($R(t)$ — множество номеров этих операций). В данный момент нужно принимать решение о распределении ресурсов только на работы фронта. Выбор некоторого распределения соответствует выбору подмножества решений (множество всех вариантов распределения в момент t определяет процесс ветвления). Определим функцию оценки $\varphi = \sum\limits_{i \in R(t)} \Delta\tau_i(t) u_i(t)$, где $\Delta\tau_i(t)$ — полный резерв времени операции $A_i \in F(t)$. Легко убедиться, что правило выполнения операций с меньшими значениями $\Delta\tau_i(t)$ в первую очередь будет соответствовать выбору решения с минимальным значением функции оценки φ. Для задачи II распределение ресурсов, удовлетворяющее этому правилу, определить нетрудно (см., например, [1]). Алгоритм распределения для задачи I описан ниже для случая $k = 1$ и $\alpha_i = 0$, $i = 1, 2, \ldots, n$ (его легко обобщить на случай $k > 1$). Недостатком большинства алгоритмов, основанных на методе ветвлений с эвристической функцией оценки, является отсутствие в общем случае оценок близости полученного решения к оптимальному. В данной работе рассмотрен частный случай $k = 1$, так как при этом удается получить оценку для времени выполнения комплекса, а также предложить некоторые приемы для улучшения решения.

В некоторых работах [7, 8] наряду с правилом распределения по величине резерва времени вводятся и другие эвристические правила. Такие правила учитывают, например, длительность операции или какие операции и в каком объеме освобождаются после выполнения данной операции и т. д.

Из других методов практический интерес представляет метод статистических испытаний с введением некоторых элементов обучения с целью увеличения вероятности появления «хороших» решений, а также метод последовательного улучшения решения путем переброски ресурсов с операций, имеющих резервы, на операции критического пути.

Иногда требуется обеспечить не минимум времени, затраченного на выполнение комплекса, а равномерное в некотором смысле использование ресурсов при заданном времени. При этом удобным является критерий вида $I = \int\limits_0^T \left(\sum u_i(t) \right)^2 dt$. Задаче с таким критерием посвящена, например, работа [9], где для решения используется физическая аналогия (I соответствует минимуму энергии некоторой физической системы).

2. Оценки времени выполнения комплекса

Будем рассматривать задачу I, считая $k = 1$ и $\alpha_i = 0$ ($i = 1, 2, \ldots, n$). Поместим на каждую операцию максимально возможное количество ресурсов $u_i = \beta_i$. При этом $\tau_i = W_i / \beta_i$ — минимальное время выполнения i-й операции.

Пусть T — время выполнения комплекса при $u_i = \beta_i$ ($i = 1, 2, \ldots, n$). Обозначим через $T_i^{\text{р}}$ самый ранний возможный момент начала

i-й операции, $T_i^{\text{п}}$ — самый поздний допустимый момент начала i-й операции, $\Delta T_i = T_i^{\text{п}} - T_i^{\text{р}}$ — полный резерв времени. Если задать времена T_i начала всех операций, так что $T_i^{\text{р}} \leqslant T_i \leqslant T_i^{\text{п}}$, и время выполнения комплекса T, то можно построить график $M(t)$ использования ресурсов во времени (график ИР). Особый интерес представляет график $M^{\text{р}}(t)$, полученный при условии, что i-я операция начинается в момент $T_i^{\text{р}}$ ($i = 1, 2, \ldots, n$), и график $M^{\text{п}}(t)$, построенный при предположении, что i-я операция начинается в момент $T_i^{\text{п}}$. На рис. 2 приведены примеры графиков $M^{\text{р}}(t)$ и $M^{\text{п}}(t)$ (пунктирная прямая показывает количество ресурсов N). Из рисунка видно, что в интервале $(0, \theta_1)$ графика $M^{\text{р}}(t)$, а также в интервале (θ_2, T) графика $M^{\text{п}}(t)$ задача распределения тривиальна. Поэтому операции (или части операций), выполняемые в этих интервалах, можно не рассматривать. В дальнейшем, не оговаривая особо, будем считать, что $M^{\text{р}}(0) > N$ и $M^{\text{п}}(T-0) > N$ ($M(t)$ считаем непрерывной справа).

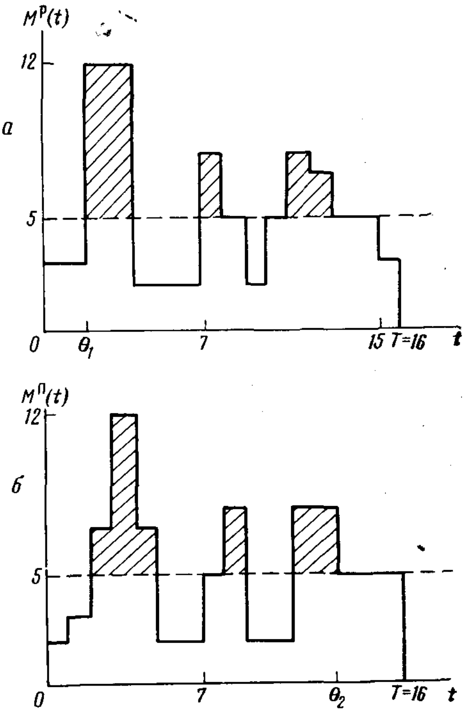


Рис. 2

Пусть ε — интервал времени, в котором $M(t) = \text{const} > N$. Положим для операций (или их частей), которые выполняются в этом интервале, $u_i(t) = \beta_i N / M(t)$ $(t \in \varepsilon)$. Тогда время выполнения этих операций $\varepsilon' = \varepsilon M(t) / N$. Проделаем такое преобразование для каждого интервала, в котором $M(t) = \text{const} > N$. Получим допустимое решение с временем выполнения комплекса

$$\theta = T + \frac{N}{S}, \tag{1}$$

где

$$S = \int_0^T (M(t) - N)\,1(M(t) - N)\,dt.$$

Определим также

$$P = \int_0^T (N - M(t))\,1(N - M(t))\,dt$$

($1(x) = 1$ при $x > 0$ и $1(x) = 0$ при $x \leqslant 0$).

Описанный способ получения допустимого решения будем называть методом пропорционального растяжения.

Заметив, что при $P = 0$ метод пропорционального растяжения дает оптимальное решение, получаем оценку

$$T + \frac{S - P}{N} \leqslant T_{\text{мин}} \leqslant T + \frac{S}{N}. \tag{2}$$

Если в качестве $M(t)$ взять $M^{\text{р}}(t)$ или $M^{\text{п}}(t)$, то можно улучшить нижнюю оценку. Для этого определим

$$C = \max_j \sum_{i=0}^{j} (S_{2i+1} - S_{2i}) \qquad (j = 1, 2, 3, \ldots), \tag{3}$$

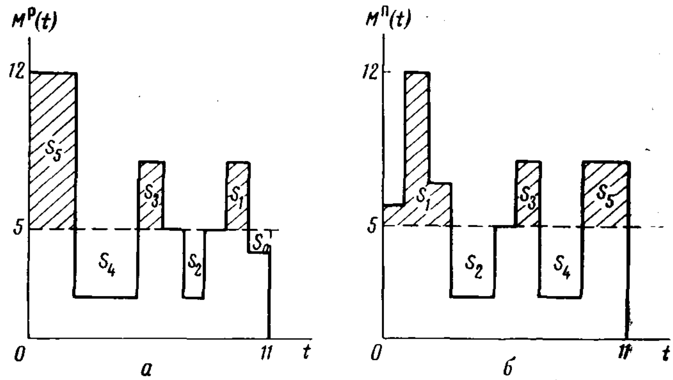
где $S_{2i+1} = (M^{\text{р}}(t) - N)\varepsilon_i$, ε_i — длина i-го интервала, в котором $M^{\text{р}}(t) = \text{const} > N$; $S_{2i} = (N - M^{\text{р}}(t))\varepsilon_j$, ε_j — длина $(j+1)$-го интервала, в котором $M^{\text{р}}(t) = \text{const} < N$, как показано на рис. 3, *а* (для графика $M^{\text{р}}(t)$) и рис. 3, *б* (для графика $M^{\text{п}}(t)$).

Теорема. Время выполнения комплекса не меньше, чем

$$T + C / N, \tag{4}$$

где T — время выполнения комплекса при времени выполнения каждой операции τ_i, а C определяется из (3).

Доказательство. Рассмотрим график $M^{\text{п}}(t)$. Предположим, что $T_{\text{мин}} < T + C / N$. Пусть $C = \sum\limits_{i=0} (S_{2i+1} - S_{2i})$ и пусть t_r — момент времени, в котором достигается этот максимум, т. е. $C = \int\limits_0^{t_r} (M(t) - N)\,dt = \int\limits_0^{t_r} M(t)\,dt - Nt_r$. За время $t_r + (T_{\text{мин}} - T)$ можно выполнить объем работ не более $Nt_r + N(T_{\text{мин}} - T) < \int\limits_0^{t_2} M(t)\,dt$. Поэтому в момент



а б

Рис. 3

$t_r + T_{\text{мин}} - T$ найдется хотя бы одна операция (или часть операции), для которой самый поздний допустимый момент начала меньше, чем t_r. Имеем для любого решения с временем выполнения комплекса θ (в том числе и оптимального)

$$\theta > t_r + (T_{\text{мин}} - T) + (T - t_r) = T_{\text{мин}}.$$

Полученное противоречие доказывает теорему. Для графика $M^{\text{р}}(t)$ доказательство получается автоматически, если изменить ход времени на обратный.

Так как $C \geqslant S - P$, то оценка (4) не хуже (2). Для графика, приведенного на рис. 3, *а*, например, имеем $S_0 = 1$, $S_1 = 3$, $S_2 = 3$, $S_3 = 3$, $S_4 = 9$, $S_5 = 14$, $S = S_1 + S_3 + S_5 = 20$, $P = S_0 + S_2 + S_4 = 13$, $C = 7 = S - P$. Следовательно, $12{,}4 = 11 + {}^7/_5 \leqslant T_{\text{мин}} \leqslant 11 + {}^{20}/_5 = 15$.

Для графика рис. 3, *б* имеем $S_0 = 0$, $S_1 = 10$, $S_2 = 6$, $S_3 = 3$, $S_4 = 6$, $S_5 = 6$, $S = S_1 + S_3 + S_5 = 19$, $P = S_0 + S_2 + S_4 = 12$, $C = 10 > S - P = 7$, т. е. $13 = 11 + {}^{10}/_5 \leqslant T_{\text{мин}} \leqslant 11 + {}^{19}/_5 = 14{,}8$.

3. Описание алгоритма

Разобьем множество номеров работ фронта на группы $R_0(t)$, $R_1(t)$, ... с одинаковыми значениями резерва $\Delta\tau_i(t)$ ($\tau_i(t) = (W_i - V_i(t))\,/\,\beta_i$, где $V_i(t)$ — объем i-й операции, выполненной к моменту t) и расположим эти группы в порядке возрастания величины резерва (в группу $R_0(t)$, например, входят критические операции фронта). Определим группу $R_p(t)$ так, чтобы

$$M_{p-1} = \sum_{i=1}^{p-1} N_i \leqslant N < \sum_{i=1}^{p} N_i = M_p,$$

где

$$N_i = \sum_{j \in R_i(t)} \beta_j.$$

Тогда распределение ресурсов по правилу первоочередного выполнения работ с меньшими значениями $\Delta\tau_i(t)$ будет иметь вид

$$u_i(t) = \begin{cases} \beta_i & \text{при } i \in R_j\ (j < p), \\ \gamma\beta_i & \text{при } i \in R_p \left(\gamma = \dfrac{N - M_{p-1}}{N_p}\right), \\ 0 & \text{при } i \in R_j\ (j > p). \end{cases}$$

Перераспределение ресурсов происходит в один из следующих моментов.

а. Резерв времени операций группы R_{p-1} становится равным резерву времени группы R_p. Это произойдет в момент $t + \delta_1$, где

$$\delta_1 = \frac{\Delta\tau(p) - \Delta\tau(p-1)}{1-\gamma}.$$

(Через $\Delta\tau(j)$ здесь обозначен резерв времени операций группы R_j.)

б. Резерв времени операций группы R_p становится равным резерву времени операций группы R_{p+1}. Это произойдет в момент $t + \delta_2$, где

$$\delta_2 = \frac{\Delta\tau(p+1) - \Delta\tau(p)}{\gamma}.$$

в. Выполнена какая-либо операция фронта. Это произойдет в момент $t + \delta_3$, где

$$\delta_3 = \min_{i \in R(t)} \frac{W_i - V_i(t)}{u_i(t)}.$$

Момент перераспределения ресурсов

$$t_1 = t + \delta,\ \text{где}\ \delta = \min\,(\delta_1, \delta_2, \delta_3).$$

Определяя $V_i(t_1) = V_i(t) + u_i(t)\delta$, $i = R(t)$ и новый фронт работ $F(t_1)$, повторяем описанный выше процесс. Таким образом, начиная с момента $t = 0$, последовательно распределяем ресурсы по всем работам.

Замечание 1. Иногда перераспределение ресурсов допустимо только в моменты окончания какой-либо операции. Этому условию легко удовлетворить, усреднив ресурсы на операциях, которые выполняются в интервале между двумя соседними моментами окончания операций, т. е. положив

$$u_i(t) = \frac{1}{t_2 - t_1} \int_{t_1}^{t_2} u_i(\tau)\,d\tau, \qquad t \in (t_1, t_2),$$

где t_1, t_2 — соседние моменты окончания операций.

Замечание 2. Если требуется, чтобы решение было целочисленным, то его нетрудно получить из нецелочисленного решения. Однако на практике проще округлять результаты до целых чисел непосредственно в процессе решения задачи.

Замечание 3. Как и все алгоритмы эвристического характера, основанные на методе ветвлений, алгоритм не требует больших вычислений. Решение на каждом шаге определяется, а число шагов равно числу операции для задачи II (так как перераспределение ресурсов происходит только в моменты окончания операций) и немногим больше в задаче I.

Таблица 1

Номер операции	1	2	3	4	5
W_i	3	1	3	1	3
β_i	3	1	3	1	3

Замечание 4. Иногда можно улучшить решение, применив следующий прием: возьмем количество ресурсов $N' > N$ и решим задачу при помощи описанного алгоритма. Полученное решение не будет в общем случае допустимым. Однако, применив метод пропорционального растяжения, получим допустимое решение, которое может оказаться лучше первоначального. Отметим, что если $T_i^p = T_i^n$ для всех операций, то описанный прием не дает улучшения.

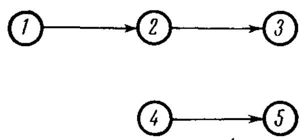


Рис. 4

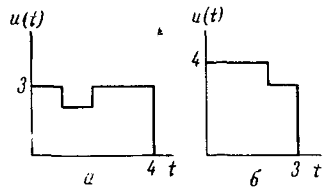


Рис. 5

Пример (рис. 4). Пусть $N = 3$ (см. табл. 1).

Применяя алгоритм, получим распределение ресурсов (рис. 5, *а*) с временем выполнения комплекса $\theta = 4$. Возьмем $N' = 4$ и снова применим алгоритм. Получим распределение ресурсов (рис. 5, *б*) с временем $\theta' = 3$. Применяя метод пропорционального растяжения, получаем допустимое решение с временем выполнения комплекса $\theta'' = 3^2/_3$.

4. Некоторые замечания

Обозначим через $K(t)$ кривую использования ресурсов для допустимого решения. Назовем «узким местом» графика $K(t)$ интервал времени, в котором $K(t) < N$. Очевидно, если график $K(t)$ не имеет узких мест, то полученное решение оптимально. Наличие узких мест соответствует возможной неоптимальности решения. Опишем эвристический прием, позволяющий в некоторых случаях устранить узкое место графика. Обозначим через Q множество операций, выполняемых в узком месте. Разобьем это множество на два подмножества: Q' и Q'' (рис. 6). Через P' обозначим множество операций, которые необходимо выполнить, чтобы начать выполнение операций множества Q', через R' — множество операций, которые можно выполнять, если выполнены операции множества Q'.

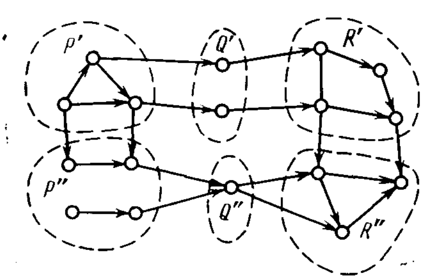


Рис. 6

Пример (рис. 7). Пусть $N = 3$ (см. табл. 2). Применяем алгоритм распределения по величине резерва времени, распределяя в первую очередь ресурсы на операции множества P'. После выполнения операций множества P' сразу начинаем выполнение операций множества Q'.

Применяя алгоритм, получаем распределение ресурсов (рис. 8, *а*) с временем $\theta = 6$. В узком месте выполняются операции A_3 и A_4. Возьмем $Q' = \{A_3\}$, $Q'' = \{A_4\}$. Тогда $P' = \{A_1\}$, $R' = \{A_5\}$. Применяя описанный прием, получаем распределение ресурсов (рис. 8, *б*) с временем выполнения комплекса $\theta' = 5{,}5$, которое является оптимальным.

Таблица 2

Номер операции	1	2	3	4	5	6
W_i	3	3	2	2	3	3
β_i	3	3	1	1	3	3

Вопрос о возможности устранения узкого места является довольно сложным и требует дальнейшего исследования. Приведем простой результат в случае, когда в узком месте выполняются две работы. Обозначим через $W(P'')$ объем работ множества P'', $W(R')$ — объем работ множества R' (рис. 6).

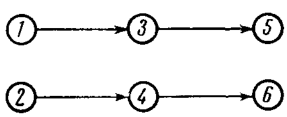


Рис. 7

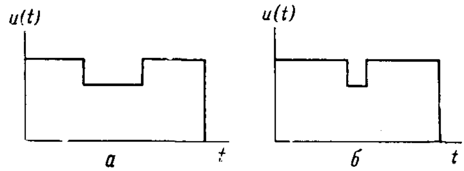


Рис. 8

Пусть $Q' = \{A_i\}$, а $Q'' = \{A_j\}$. Тогда необходимым условием устранимости узкого места Δ при выбранном разбиении Q' и Q'' является

$$W(P'') \geqslant (N - \beta_i)\Delta, \quad W(R') \geqslant (N - \beta_j)\Delta.$$

Таким образом, если $\Delta > \min\left(\frac{W(P'')}{N - \beta_i}, \frac{W(R')}{N - \beta_j}\right)$, то узкое место неустранимо.

Для предыдущего примера имеем $W(P'') = 3$, $W(R') = 3$, $\beta_i = \beta_j = 1$, $N = 3$, $\Delta = 2$:

$$\Delta = 2 > \min\left(\frac{3}{3-1}, \frac{3}{3-1}\right) = 1{,}5.$$

Действительно, в оптимальном решении остается узкое место $\Delta' = 0{,}5$.

Таким образом, алгоритм распределения по величине резерва времени не столько позволяет получить оптимальное решение, сколько дает информацию о структуре сети (наличие узких мест, их площадь и т. д.). Дальнейшие исследования предполагается вести в направлении решения вопроса об устранимости узких мест.

Поступила в редакцию
10 декабря 1965 г.

ЦИТИРОВАННАЯ ЛИТЕРАТУРА

1. Moder J. J., Phillips C. R. Project Management with CRM and PERT. Reinhold Publishing Corporation, 1964.
2. Hu T. C. Parallel Sequencing and Assembly Line Problems Opus Res. v. 9, No. 6, 1961.
3. Lomnicki Z. A. A «brauch — and — bound» Algorithm for the Rxact Solution of the Three-Machine Scheduling Problem. Operat. Res. Quart., v. 16, No. 1, 1965.

4. Broohs G. H., White C. R. An Algorithm for Finding Optimal Solutions to the Production Scheduling Problem. J. Industr. Engng, v. 16, No. 1, 1965.
5. Шкурба В. В. Вычислительные схемы решения задач теории расписаний. Кибернетика, № 3, 1965.
6. Iohn D. C. Little, Katta G. Murty, Dura W. Sweeney and Caroline Karel. An Algorithm for the Traveling Salesman Problem. Opns Res., v. 11, No. 6, 1963.
7. Schwartz F. C. An Automatic Sequencing Procedure with Application to Parallel Programming. J. Assoc. Compt. Machinery, v. 8, No. 4, 1961.
8. Wiest I. D. Some Properties of Schedules for Large Projects with Limited Resources. Opns. Res., v. 12, No. 3, 1964.
9. Разумихин Б. С. Задача об оптимальном распределении ресурсов. Автоматика и телемеханика, т. XXVI, № 1, 1965.

HEURISTIC APPROACH TO THE SOLUTION OF DYNAMIC PROBLEMS OF RESOURCES DISTRIBUTION

V. N. BURKOV, S. E. LOVETSKI

The dynamic problem of resources distributions over a set of activities (operations) in the network planning systems are considered. It is assumed that the resources, after using in one activity, can be used for carrying out other activities of the given class (for instance, machines, equipment, people, and so on).

An heuristic approach to the solution of dynamic problems is suggested. In a number of cases the algorithm gives an optimal solution. Optimality degree estimates have been obtained.
