Эвристический алгоритм определения последовательности работ

Д.А. Калинов
(*Институт проблем управления РАН, Москва*)

Введение

В данной статье рассматривается одна из задач теории расписаний в достаточно простой постановке. Приводится приближенный алгоритм решения задачи и исследуется эффективность предложенного алгоритма.

1. Постановка задачи и метод ее решения

Пусть имеется один исполнитель. Требуется выполнить *n* работ с продолжительностью $t_i, i=\overline{1,n}$. Заданы продолжительности работ $l_i, i=\overline{1,n}$ и сроки выполнения $d_i, i=\overline{1,n}$. Требуется определить последовательность выполнения работ, минимизирующую сумму просрочек работ.

Для простоты предположим, что все величины целочисленные. Очевидно, что работы целесообразно планировать целиком (не допуская разрывов в выполнении любой работы). В случае одного исполнителя разрывы не позволяют перераспределить время выполнения работ так, чтобы добиться преимущества перед неразрывным планированием.

Для решения задачи предлагается использование эвристического алгоритма, не гарантирующего наилучшего результата, но имеющего приемлемую вычислительную сложность.

В алгоритме используется метод ветвления, планирование осуществляется по шагам времени. На каждом шаге для каждой незапланированной работы будет вычисляться индекс – вещественное число (“ценность” работы), которое будет определять, какую из работ следует планировать на рассматриваемом шаге.

Дальнейшее описание потребует введения обозначений. Способ вычисления ряда величин будет приведен в описании алгоритма решения задачи.

Введем понятие “текущего времени” или “времени шага планирования”, обозначаемого t – это суммарная длительность уже запланированных работ (на каком-то шаге). Также обозначим $N=\{1,\ldots,n\}$ – множество номеров всех работ, $P_\kappa \subseteq N, \kappa=\overline{1,n}$ – множество номеров запланированных работ на κ-м шаге планирования; соответственно, $U_k = N \setminus P_k$ – множество номеров еще не запланированных работ на k-м шаге планирования. Множеством работ-кандидатов на κ-м шаге обозначим $C_k \subseteq U_k$.

Резервом работы на k-м шаге планирования назовем величину $r_i=[d_i-(t+l_i)]_+, i\in U_k$, где t – время k-го шага планирования ($t=\sum_{i\in P_k} l_i$).

Суммой гарантированных потерь (первого уровня) i-й работы обозначим величину s_i.

Индексом i-й работы обозначим величину $ind_i = w\cdot s_i + (1-w)\cdot r_i, i=\overline{1,n}$. Соответственно, $w\in[0,1]$ – весовой коэффициент.

Максимальную длительность незапланированной работы (для некоторого шага) обозначим $l_{\max}$ ($l_{\max}=\max\{l_i, i\in U_k\}$, для k-го шага). Минимальный срок выполнения среди незапланированных работ обозначим d_{min}.

Получаемый план для k-го шага обозначим p_k (номер работы, которую следует выполнять k-й).

Последовательность действий для решения задачи:

1. Шаг 1: k=1. Время шага t=0. При этом $U_1 = N, P_1=\varnothing$.
2. Вычисляем максимальную длительность незапланированных работ: $l_{\max}=\max\{l_i, i\in U_k\}$.

3. Определяем множество кандидатов C_k шага k среди незапланированных работ: $i \in C_k$, если $t + l_i + l_{\max} > d_i$, $i \in U_k$.

4. Если кандидатов нет ($C_k = \varnothing$), то переходим к пункту 9.

5. Вычисляем значение резерва и СГП для каждой работы-кандидата: $r_i = [d_i - (t + l_i)]_+$ и $s_i = [(t + l_i) - d_i]_+ + \sum_{j \in (C_k \setminus \{i\})} [(t + l_i + l_j) - d_j]_+$, $i \in C_k$.

6. Вычисляем значение индекса для каждой работы-кандидата: $ind_i = w \cdot s_i + (1 - w) \cdot r_i$, $i \in C_k$.

7. Выбираем работу: $p_k = \arg\min\{ind_i, i \in C_k\}$ (если индексы равны, можно выбрать любую работу с минимальным значением).

8. Исключаем запланированную работу из незапланированных работ для следующего шага: $U_{k+1} = U_k \setminus \{p_k\}$. Увеличиваем время следующего шага на длительность запланированной работы: $t := t + l_{p_k}$. Переходим к пункту 15.

9. Если кандидатов нет, то находим минимальный крайний срок еще не запланированной работы и сдвигаем время шага на эту величину.

10. Снова осуществляем поиск кандидатов: $i \in C_k$ *если* $t + l_i + l_{\max} > d_i, i \in U_k$. Заметим, что среди кандидатов будет хотя бы одна работа

11. Вычисляем значение резерва и СГП для каждой работы-кандидата: $r_i = [d_i - (t - d_{\min} + l_i)]_+$ и $s_i = [(t + l_i) - d_i]_+ + \sum_{j \in (C_k \setminus \{i\})} [(t + l_i + l_j) - d_j]_+$, $i \in C_k$. При этом следует обратить внимание, что резерв считается по "старому" времени шага.

12. Вычисляем значение индекса для каждой работы-кандидата: $ind_i = w \cdot s_i + (1 - w) \cdot r_i$, $i \in C_k$.

13. Выбираем работу: $p_k = \arg\min\{ind_i, i \in C_k\}$ (если индексы равны, можно выбрать любую работу).

14. Исключаем запланированную работу из незапланированных работ для следующего шага: $U_{k+1} = U_k \setminus \{p_k\}$. Возвращаем время шага и увеличиваем его на длительность запланированной работы: $t := t - d_{\min} + l_{p_k}$.

15. Если запланировано меньше, чем n-1 работа, то выполняем следующий шаг: $k := k + 1$, переходим к пункту 2. Иначе переходим к следующему пункту.

16. На последнем шаге планируем оставшуюся незапланированной последнюю работу.

Оценка вычислительной сложности представленного алгоритма: в среднем $\frac{7}{8} \cdot n^2 + \frac{1}{16} \cdot n^3 \approx n^3$ действий.

Приведенный алгоритм не гарантирует получение оптимального решения. Это демонстрирует следующий пример: Пусть n=8, l_1=1, l_2=2, …, l_8=8, d_1=d_2=…=d_8=25. Определить оптимальное решение в этом случае легко: {1,2,3,…,8} (короткие работы следует планировать раньше). При этом получаем результат $F_{опт}$=14. Алгоритм предлагает следующий вариант: {1,2,3,4,8,5,6,7}, что приводит к результату $F_{алг}$=15. Более детальный анализ эффективности алгоритма приведен в следующем разделе.

2. Статистический анализ

Метод статистического анализа

Для оценки эффективности предложенных алгоритмов использовался метод Монте-Карло со следующими параметрами: для 8 работ создавались псевдослучайные начальные данные (длительности и крайние сроки выполнения), получаемые по определенному набору распределений. Эти наборы распределений названы ситуацией. Для каждой ситуации рассматривалось по 1000 вариантов начальных условий. Ситуации отличаются друг от друга степенью "напряженности" варианта начальных условий (возможностью успеть выполнить все работы вовремя).

Для каждого варианта методом перебора находилось лучшее решение ($F_{опт}$), худшее ($F_{худ}$), и среднее значение результата по всем возможным вариантам ($F_{сред}$). Последняя величина представляет собой математическое ожидание результата, если план выбирается равновероятно из всех возможных вариантов. Для решения алгоритма ($F_{алг}$) представляется целесообразным вычислить два коэффициента: $\%_{к\,худ} = \dfrac{F_{алг} - F_{опт}}{F_{худ} - F_{опт}}$ и $\%_{к\,сред} = \dfrac{F_{алг} - F_{опт}}{F_{сред} - F_{опт}}$. Объяснение желательности использования обоих коэффициентов состоит в следующем: в общем случае (и как правило) $F_{худ} - F_{опт} \neq 2 \cdot (F_{сред} - F_{опт})$. Это иллюстрируют следующие графики, полученные для ситуации **С1** (описание параметров распределений начальных условий для этой ситуации указано ниже).

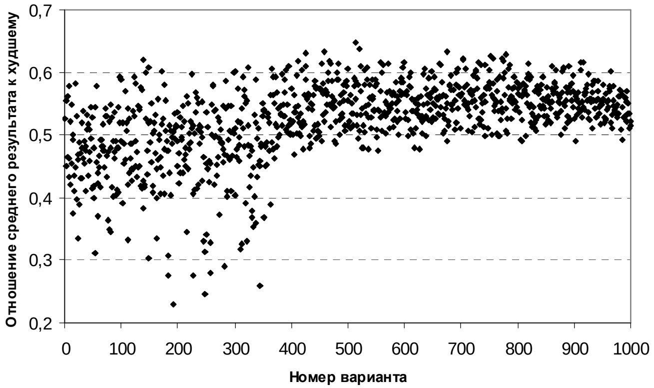

Рис. 1.

На рисунке 1 по оси абсцисс отложены номера вариантов, упорядоченных по возрастанию оптимального результата, по оси ординат – отношение $\dfrac{F_{сред} - F_{опт}}{F_{худ} - F_{опт}}$.

Результаты анализа при различных условиях

Ситуация С1. Использовались следующие распределения: длительности работ – нормальное распределение с математическим ожиданием 10 и среднеквадратичным отклонением 2,5. Крайние сроки – равномерное распределение от 1 до 120. Эта ситуация моделирует в среднем легкую напряженность, так как математическое ожидание длительности всех 8 работ равно 80.

Получены следующие результаты:

	$\%_{к\,сред}$	$\%_{к\,худ}$
Среднее значение	0,164	0,089
Стандартное отклонение	0,768	0,419
Минимальное значение	0,000	0,000
Максимальное значение	9,864	5,714
% оптимальных решений	93,2	

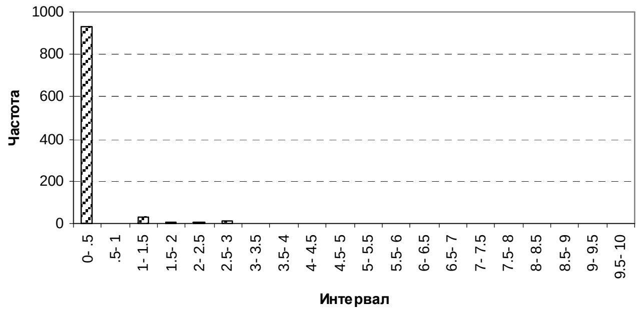

Рис. 2. Распределение значений величины $\%_{к\,сред}$ для ситуации С1

Распределение значений величины $\%_{к\,худ}$ практически не отличается от распределения величины $\%_{к\,сред}$, поэтому дополнительно не приводится.

Затем проводились исследования зависимости величин $\%_{к\,худ}$ и $\%_{к\,сред}$ от используемого весового коэффициента w (напомним, что в алгоритме индекс рассчитывался следующим образом: $ind_i = w \cdot s_i + (1-w) \cdot r_i,\ i \in C_k$). Получены следующие результаты:

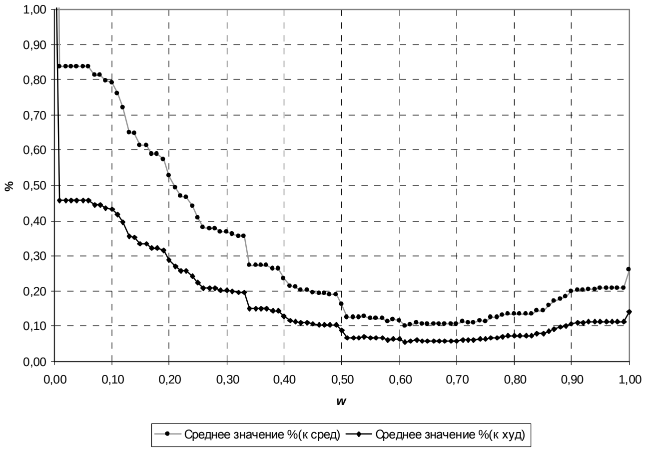

Рис. 3. Зависимость средних значений $\%_{к\ худ}$ и $\%_{к\ сред}$ от весового коэффициента **w**

Наилучшее значение *w* равно 0,61 (исследование проводилось с шагом 0,01), причем это значение *w* дает минимум обоих рядов. Обнаруженная зависимость показывает правильность включения в индекс и СГП, и резерва. Более того, можно определять наилучшее значение весового коэффициента между этими величинами, улучшая тем самым среднестатистическую эффективность алгоритма.

Ситуация С2. Распределения были изменены в сторону увеличения “напряженности” вариантов: длительности работ – нормальное распределение с математическим ожиданием 10 и среднеквадратичным отклонением 2,5. Крайние сроки – равномерное распределение от 1 до 80. Эта ситуация моделирует среднюю напряженность. Получены следующие результаты:

	$\%_{к\ сред}$	$\%_{к\ худ}$
Среднее	0,774	0,420
Стандартное отклонение	2,400	1,257
Минимум	0,000	0,000
Максимум	22,071	12,195
% оптимальных решений	79,3	

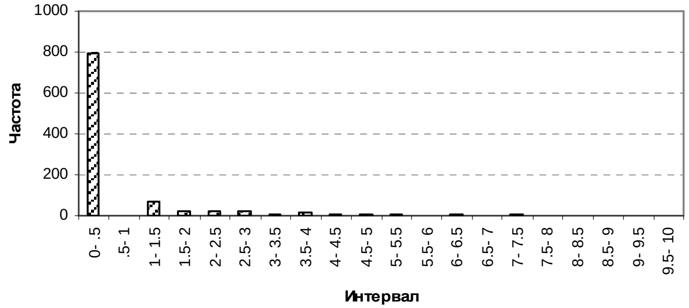

Рис. 4. Распределение значений величины $\%_{к\ сред}$ для ситуации С2

Результаты зависимости $\%_{к\ худ}$ и $\%_{к\ сред}$ от используемого весового коэффициента *w*:

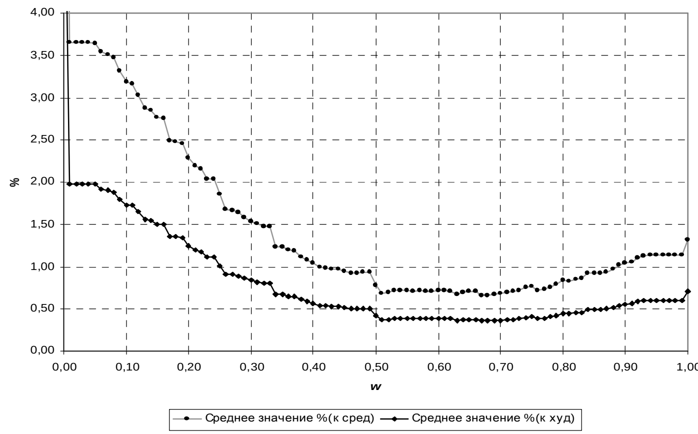

Рис. 5. Зависимость средних значений $\%_{к\ худ}$ и $\%_{к\ сред}$ от весового коэффициента w

Наилучшее значение w равно 0,67 и 0,68 (шаг исследования 0,01), эти значения *w* дают минимум обоих рядов.

Ситуация С3

Распределения были изменены до значительной “напряженности” вариантов: длительности работ – нормальное распределение с математическим ожиданием 10 и среднеквадратичным отклонением 2,5. Крайние сроки – равномерное распределение от 1 до 40. В такой ситуации вероятность уложиться в установленные крайние сроки выполнения крайне мала.

Алгоритм принес следующие результаты:

	$\%_{к\ сред}$	$\%_{к\ худ}$
Среднее	2,325	1,171
Стандартное отклонение	5,837	2,990
Минимум	0,000	0,000
Максимум	59,097	31,034
% оптимальных решений	69,2	

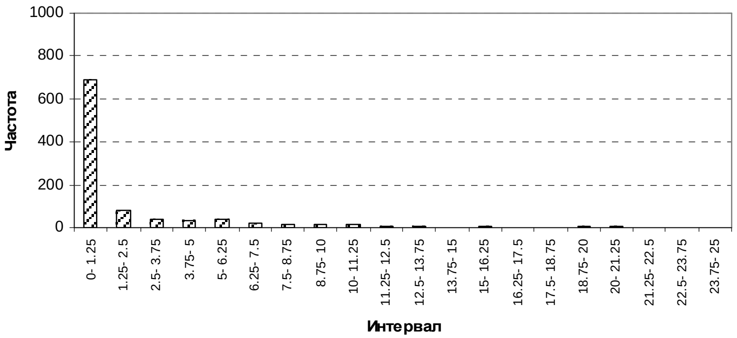

Рис. 6. Распределение значений величины $\%_{к\ сред}$ для ситуации С3

Результаты зависимости $\%_{к\ худ}$ и $\%_{к\ сред}$ от используемого весового коэффициента w:

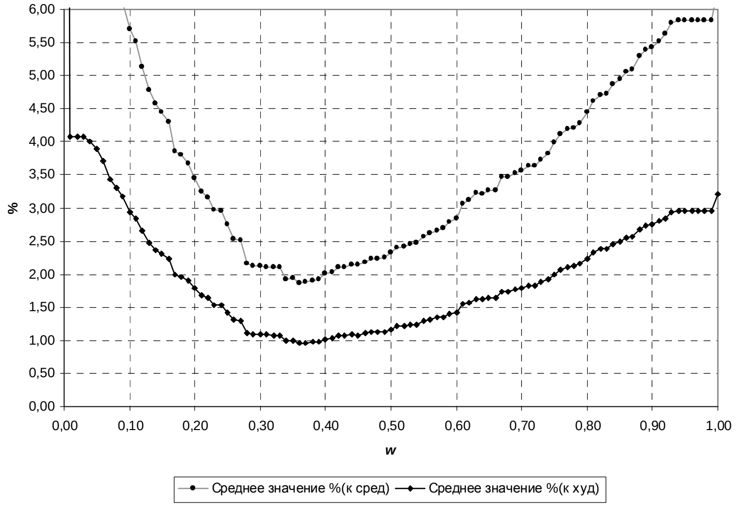

Рис. 7. Зависимость средних значений $\%_{к\ худ}$ и $\%_{к\ сред}$ от весового коэффициента w

Наилучшее значение *w* равно 0,36, причем это значение *w* дает минимум обоих рядов. Любопытно отметить, что с ростом напряженности варианта лучшее значение *w* смещается влево (это также подтверждается результатами статистического анализа, которые не отображены в настоящей работе).

Выводы

Основываясь на результатах статистического анализа можно сделать основной вывод: алгоритм показал достаточно высокую эффективность и может быть использован для решения практических задач. В ходе статистического анализа было проверено более 7000 вариантов начальных данных, при этом более чем в 65% случаев эвристический алгоритм приводил к оптимальному результату. Особенно эффективным алгоритм показал себя для наборов начальных данных с низкой напряженностью: обнаружен высокий процент оптимальных решений.

В ходе анализа зависимости среднестатистической эффективности алгоритма от параметра (весового коэффициента w) сделан следующий вывод: изменение w может заметно увеличить среднестатистическую эффективность алгоритма. При этом можно дать следующую общую рекомендацию: при низкой напряженности целесообразно использовать значение $w \approx 0{,}7 - 0{,}8$, при высокой – $w \approx 0{,}4$.

Заключение

Следует отметить, что представленные результаты работы сами по себе представляют достаточно невысокую практическую ценность, так как рассмотренная в работе задача является слишком узкой. Целью дальнейших исследований является применение использованного подхода на более сложные постановки задач.

Литература

1. Бурков В. Н., Горгидзе И. А., Ловецкий С. Е. *Прикладные задачи теории графов*. – Тбилиси: Мецниереба, 1974.
2. Бурков В. Н., Заложнев А. Ю., Новиков Д. А. *Теория графов в управлении организационными системами*. М.: СИНТЕГ, 2001.
3. Голенко Д. И. *Статистические методы сетевого планирования и управления*. М.: Наука, 1968.