УДК 681.3
Э.И. Ватутин, канд. техн. наук, доцент, кафедра вычислительной техники, ЮЗГУ (e-mail: evatutin@rambler.ru)
А.С. Романченко, канд. техн. наук, доцент, начальник учебно-методического управления, ЮЗГУ (e-mail: romanchas@rambler.ru)
В.С. Титов, д-р техн. наук, профессор, зав. кафедрой вычислительной техники, ЮЗГУ (e-mail: titov-kstu@rambler.ru)

# ИССЛЕДОВАНИЕ ВЛИЯНИЯ ПОРЯДКА РАССМОТРЕНИЯ ПАР НА КАЧЕСТВО РАСПИСАНИЙ ПРИ ИСПОЛЬЗОВАНИИ ЖАДНОГО ПОДХОДА

*Приведено описание метода построения расписаний с использованием жадной стратегии выбора пар и его модификация, отличающаяся порядком рассмотрения пар. Показано, что различный порядок рассмотрения пар при использовании жадной стратегии приводит к существенно различному качеству оптимизации как частных, так и интегрального показателя качества. Рассмотренный подход к нахождению квазиоптимального решения хорошо поддается распараллеливанию и может быть эффективно реализован с использованием грид-систем.*

*Ключевые слова: дискретная комбинаторная оптимизация, расписания, грид-системы*

***

Одной из задач дискретной комбинаторной оптимизации является задача составления различных расписаний (планирование запуска работ или задач, движение транспорта, графики работ). Задача составления расписаний занятий учебных заведений (школы, университета и т.д.), как известно, является одной из наиболее сложных в своем классе, т.к. требуется вести учет большого числа ограничений, зачастую взаимоисключающих друг друга, и частных критериев качества [1]. Для составления расписаний известно достаточно большое количество подходов, основанных как на общих положения комбинаторной оптимизации (различные виды перебора, жадные алгоритмы, генетические алгоритмы, имитация отжига, муравьиные алгоритмы и т.д.), так и учитывающих специфику задачи в ее конкретной постановке. В данной работе приведено описание результатов, полученных в ходе применения жадной стратегии в задаче составления расписания занятий университета. Известно, что данная стратегия далеко не всегда приводит к получению решений высокого качества, однако она имеет право на жизнь как довольно простая как с точки зрения трудоемкости реализации, так и по необходимым вычислительным затратам. Полученные с ее использованием результаты могут являться отправной точкой для анализа качества решений, полученных с использованием более сложных подходов, качество которых должно быть точно не ниже жадного.

Исходными данными для решения задачи является прежде всего множество пар $\Theta = \{p_i\}, i = \overline{1, N_\Theta}$, каждая из которых в первом приближении представляет собой кортеж $p_i = \langle t(p_i), G(p_i), n(p_i), a(p_i), v(p_i) \rangle$, где $t(p_i) \in T$ – преподаватель, проводящий занятие; $T = \{t_i\}, i = \overline{1, N_T}$ –

множество всех преподавателей вуза; $G(p_i) \subset G$ – множество учебных групп, объединенных в поток и присутствующих на проводимом занятии (в простейшем случае $|G(p_i)| = 1$, т.е. на занятии присутствует лишь одна группа); $G = \{g_i\}, i = \overline{1, N_G}$ – множество учебных групп университета; $n(p_i)$ – длительность учебного занятия в академических часах (1 – для занятий, читающихся один раз в две недели; 2 – для занятий, читающихся на регулярной основе каждую неделю); $a(p_i) \in A$ – аудитория, в которой запланировано проведение учебного занятия; $A = \{a_i\}, i = \overline{1, N_A}$ – множество учебных аудиторий (аудиторный фонд); $v(p_i) = \left\langle v^{(d)}(p_i), v^{(t)}(p_i) \right\rangle \in V$ – время проведения занятия (выбирается в виде дискретных значений дня недели $v^{(d)}(p_i) \in \{1, 2, ..., N_D\}$, $N_D = 7$ – число рабочих дней, и номера пары $v^{(t)}(p_i) \in \{1, 2, ..., N_P\}$, $N_P = 8$ – теоретически максимальное число пар в день, с учетом номера недели в двухнедельном цикле); $V = \{v_i\}, i = \overline{1, N_V}$ – множество временны□х позиций, в которые может быть установлена пара (7 рабочих дней по 8 пар с учетом двухнедельного цикла). Кроме того, в указанные кортежи могут быть включены дополнительные компоненты (например, пожелания преподавателей по аудиториям, учебным корпусам и времени проведения пары, тип пары (лекционное, лабораторное или практическое занятие) и т.д.). Учебные аудитории могут находиться в различных учебных корпусах $l_i \in L, i = \overline{1, N_L}$.

Множество пар формируется отталкиваясь от текущей учебной нагрузки университета на семестр, при этом должны быть априорно известны состав множества $P$ и значения $t(p_i)$, $G(p_i)$ и $n(p_i)$ входящих в него пар $p_i$. Искомыми являются значения $a(p_i)$ и $v(p_i)$. Размерность решаемых на практике задач определяется прежде всего размерностью исходных данных $N_T \simeq 100 \div 1000$, $N_G \simeq 100$, $N_A \simeq 100$, $N_V \simeq 100$ и числом пар в расписании $N_\Theta \simeq 1000$, что не позволяет отыскание оптимального расписания за приемлемое время, т.к. при полном переборе для этого потребуется построение порядка $O(N_P! \cdot N_V N_A)$ решений, поэтому для поиска решений необходимо использовать различные эвристические подходы.

При построении расписания необходим учет ряда ограничений. Прежде всего необходим учет «накладок», т.е. пересечений различных пар по преподавателю (преподаватель не может читать две пары одновременно), группе (группа не может одновременно находиться на различных занятиях) или аудитории (в одной аудитории не может проводиться более одной пары одновременно) в одно время:

$$\overline{\exists} p_i, p_j, i, j = \overline{1, N_P},\ i \neq j:$$
$$\left[\left(t(p_i) = t(p_j)\right) \vee \left(G(p_i) \cap G(p_j) \neq \varnothing\right) \vee \left(a(p_i) = a(p_j)\right)\right] \wedge \qquad (1)$$
$$\wedge \left(v(p_i) \cap v(p_j) \neq \varnothing\right).$$

Существуют ограничения на максимальное (обычно 3–4) и минимальное (обычно 1–2) число пар в день для преподавателя и учебной группы:

$$\forall v_i^{(d)}, t_j, i = \overline{1, N_D}, j = \overline{1, N_T}: p_{\min}^{(t)} \leq \left|P\left(v_i^{(d)}, t_j\right)\right| \leq p_{\max}^{(t)}, \qquad (2)$$

$$\forall v_i^{(d)}, g_j, i = \overline{1, N_D}, j = \overline{1, N_G}: p_{\min}^{(s)} \leq \left|P\left(v_i^{(d)}, g_j\right)\right| \leq p_{\max}^{(s)}, \qquad (3)$$

где $P\left(v_i^{(d)}, t_j\right) = \left\{p_{j_1}, p_{j_2}, ..., p_{j_n}\right\} \subset P$ – множество пар, установленных в день $v_i^{(d)}$ для преподавателя $t_j$, причем $p_{j_k} \in P\left(v_i^{(d)}, t_j\right) \Leftrightarrow \left(v^{(d)}\left(p_{j_k}\right) = v_i^{(d)}\right) \wedge \left(t\left(p_{j_k}\right) = t_j\right)$;
$P\left(v_i^{(d)}, g_j\right) = \left\{p_{j_1}, p_{j_2}, ..., p_{j_m}\right\} \subset P$ – множество пар, установленных в день $v_i^{(d)}$ для группы $g_j$, причем $p_{j_k} \in P\left(v_i^{(d)}, g_j\right) \Leftrightarrow \left(v^{(d)}\left(p_{j_k}\right) = v_i^{(d)}\right) \wedge \left(g_j \in G\left(p_{j_k}\right)\right)$.
Обычно ограничение на максимальное число пар в день для преподавателей, в расписании которых допускаются «окна», записывается жестче:

$$\underbrace{\max_k v^{(t)}\left(p_{j_k}\right) - \min_k v^{(t)}\left(p_{j_k}\right) + 1}_{\textit{число пар в день с учетом времени их установки}} \leq p_{\max}^{(t)}, p_{j_k} \in P\left(v_i^{(d)}, t_j\right), \qquad (4)$$

т.е. производится расчет числа пар в день с учетом «окон» между ними. На рабочее время преподавателя и на время обучения группы также возможны ограничения, фактически реализуемые путем исключения из множества $V$ некоторых значений (например, группы очной формы обучения не должны иметь пар в воскресенье, обучение групп производится по сменам, преподаватель может быть задействован в регулярно проводимых мероприятиях вуза и т.д.).

В расписании студентов «окна» обычно недопустимы:

$$\underbrace{\max_k v^{(t)}\left(p_{j_k}\right) - \min_k v^{(t)}\left(p_{j_k}\right) + 1}_{\textit{число пар в день с учетом времени их установки}} - \underbrace{\left|P\left(v_i^{(d)}, g_j\right)\right|}_{\textit{число пар в день}} = 0, p_{j_k} \in P\left(v_i^{(d)}, g_j\right). \qquad (5)$$

Кроме того, обычно на практике вводится ряд дополнительных ограничений: постановка некоторых пар строго подряд и/или в строго заданную аудиторию (например, продолжительные лабораторные работы, требующие спецоборудования), постановка некоторых пар последними в выбранном учебном дне (например, физкультура), запрет на перемещение групп между корпусами в течение рабочего дня, ограничение на число выходных дней студентов и т.д.

Ограничения (1) и (5) являются наиболее жесткими, в то время как ограничения (2–4) могут быть нарушены в крайнем случае, например, при взаимной противоречивости ограничений и невозможности построения расписания без ослабления ограничений.

Оценка качества результирующего расписания [2] производится с использованием штрафной функции, фактически представляющей собой сумму взвешенных нормированных значений частных показателей качества:

$$
\begin{aligned}
f(\Theta) = {} & \frac{K_w^{(t)} t_w(\Theta)}{2N_D w_{\max}^{(t)} T} + \frac{K_w^{(s)} s_w(\Theta)}{2N_D w_{\max}^{(s)} N_G} + \frac{K_{wt}^{(t)} t_{wt}(\Theta)}{2N_D p_{\max}^{(t)} N_T} + \frac{K_{wt}^{(s)} s_{wt}(\Theta)}{2N_D p_{\max}^{(s)} N_G} + \\
& + \frac{K_{wt}^{(a)} a_{wt}(\Theta)}{2N_D N_P N_A} + \frac{K_a^{(w)} w_a(\Theta)}{|\Theta|} + \frac{K_{gr}^{(w)} w_{gr}(\Theta)}{|\Theta|} + \frac{K_{gr}^{(t)} t_{gr}(\Theta)}{2N_D N_T} + \\
& + \frac{K_{p_{\min}}^{(t)} t_{p_{\min}}(\Theta)}{2N_D N_T} + \frac{K_{p_{\max}}^{(t)} t_{p_{\max}}(\Theta)}{2N_D N_T} + \frac{K_{l_{\max}}^{(t)} t_{l_{\max}}(\Theta)}{2N_D N_T} + \\
& + \frac{K_{p_{\min}}^{(s)} s_{p_{\min}}(\Theta)}{2N_D N_G} + \frac{K_{p_{\max}}^{(s)} s_{p_{\max}}(\Theta)}{2N_D N_G} + \frac{K_{l_{\max}}^{(s)} s_{l_{\max}}(\Theta)}{2N_D N_G} + \\
& + \frac{K_{ed}^{(g)} s_{ed}}{2N_D N_G} + \frac{K_m^{(t)} t_m}{2N_D N_T \left(\max m_{ij} + 1\right)} + \frac{K_m^{(g)} s_m}{2N_D N_G \left(\max m_{ij} + 1\right)} + \\
& + \frac{K_{lw}^{(t)} t_{lw}(\Theta)}{2N_D N_T \cdot \left(\max m_{ij} + 1\right)} + \frac{K_{wa}^{(p)} p_{wa}(\Theta)}{|\Theta|},
\end{aligned}
\tag{6}
$$

где $K_i^{(j)}$ – соответствующий весовой коэффициент (индекс $t$ – обозначение преподавателя, $s$ – группы, $a$ – аудитории, $w$ – элемента учебной нагрузки, $wt$ – таблицы рабочего времени, $gr$ – пожелания по группировке учебной нагрузки, $ed$ – числа свободных дней, $m$ – числа перемещений между корпусами, $l_{\max}$ – максимального числа лекций в день, $p_{\min}$ и $p_{\max}$ – минимального и максимального числа пар в день, $lw$ – число нарушений пожеланий по корпусу); $t_x(\Theta)$, $s_x(\Theta)$, $a_x(\Theta)$, $w_x(\Theta)$, $p_x(\Theta)$ – выбранный частный показатель качества для преподавателя, группы, аудитории, элемента нагрузки или пары соответственно; $m_{ij}$ – элемент матрицы удаленности корпусов, равный условной «дистанции» между корпусами. Лучшему расписанию соответствует меньшее значение штрафной функции, коэффициенты которой выбираются экспертным путем. Оценка качества расписания является довольно трудоемкой операцией [1], требующей реализации подсистемы кэширования частных показателей качества с целью избежания их повторного пересчета.

Жадная стратегия составления расписания заключается в следующем. Из множества еще не рассмотренных пар $\Theta^*$ выбирается очередная пара $p_i \in \Theta^*$ и определяется множество допустимых позиций $V^+(p_i) = \{v_1(p_i), v_2(p_i), ..., v_K(p_i)\}$, в которых она может быть размещена.

Для каждой из позиций $v_j(p_i) \in V^+(p_i)$ производится расчет приращения штрафной функции

$$\Delta f\left(v_j(p_i)\right) = f\left(\left(v(p_i) \to v_j\right) \in \Theta\right) - f(\Theta), \tag{7}$$

и выбирается позиция $v_{j^+}(p_i)$ с минимальным приращением штрафной функции, где $j^+ = \arg\min_j \Delta f\left(v_j(p_i)\right)$. При этом для каждой позиции производится поиск лучшей аудитории (перебор всех возможных аудиторий повышает необходимые вычислительные затраты на несколько порядков). Далее пара размещается в выбранной позиции $v_{j^+}(p_i)$ и исключается из множества нерассмотренных пар: $\Theta^* := \Theta^* \setminus \{p_i\}$. Описанный процесс повторяется до тех пор, пока не будут рассмотрены все пары $\left(\Theta^* = \varnothing\right)$ – расписание составлено полностью, либо пока в множестве нерассмотренных пар останутся пары, которые невозможно установить из-за нарушения ограничений (в таком случае в штрафную функцию целесообразно добавить слагаемое $+\dfrac{K^{(\Theta^*)}\left|\Theta^*\right|}{\left|\Theta\right|}$, благодаря которому из двух расписаний с прочими равными частными критериями лучшим будет считаться то, у которого большее число пар установлены в расписание).

При подобном жадном составлении расписания важен порядок рассмотрения пар из множества $\Theta^*$, при этом пары можно рассматривать в произвольном порядке или же вводить специальные требования к порядку их рассмотрения.

При рассмотрении пар в произвольном порядке производится формирование массива номеров пар $[1, 2, 3, ..., N_\Theta]$ с последующим перемешиванием его компонентов (например, $[1, 2, 3, 4, 5] \to [2, 4, 1, 5, 3]$), т.е. фактически получением различных упорядоченных $r$-выборок без повторения (перестановок). Целью не является получение всех $N_\Theta!$ перестановок (что практически невозможно для больших $N_\Theta$), а лишь малой их части (несколько сотен – тысяч перестановок) с целью дальнейшего построения группы расписаний, оценки их качества и выбора наилучшего или группы из $n$ лучших с целью дальнейшего анализа коллективом экспертов. Простейшим алгоритмом псевдослучайного перемешивания элементов массива является поочередный просмотр его элементов и обмен местами $i$-го элемента с $r_i$-м, где $r_i$ – $i$-е псевдослучайное число, полученное с использованием генератора псевдослучайных чисел (ГПСЧ). Задавая различные начальные значения $r_0$ для ГПСЧ (англ. rand seed), можно получать различные перестановки, вероятность дублирования которых довольно мала при условии $n \ll N_\Theta!$. Например:

$$r_0 = 0{:}\,[2, 3, 5, 1, 4],$$
$$r_0 = 1{:}\,[1, 4, 5, 2, 3],$$
$$r_0 = 2{:}\,[5, 2, 3, 4, 1],$$
$$r_0 = 3{:}\,[1, 3, 2, 5, 4],$$
$$r_0 = 4{:}\,[1, 3, 4, 2, 5].$$

При подобном подходе процесс построения различных расписаний тривиально распараллеливается на *n* не зависящих друг от друга подзадач, не связанных между собой обменами данных, что позволяет организацию эффективного параллельного исполнения с использованием различных параллельных вычислительных систем (многоядерные/многопроцессорные машины, кластеры, суперкомпьютеры, грид).

Используя указанную особенность и отталкиваясь от реальных данных (весенний семестр 2012 учебного года ЮЗГУ) был организован вычислительный эксперимент, целью которого было исследования влияния порядка рассмотрения пар при использования жадной стратегии построения расписания. Всего в ходе данного вычислительного эксперимента было построено 30 расписаний, количественная оценка качества которых приведена в таблице (приведены значения наиболее важных показателей качества для 5 лучших и 5 худших решений). Вычислительные затраты на выполнение расчетов составили приблизительно 10 часов машинного времени одного ядра процессора Core 2 Duo 6300 1,86 ГГц.

Таблица. Результаты вычислительного эксперимента

| № | $t_w$ | $s_w$ | $t_m$ | $s_m$ | $t_{p_{\min}}$ | $t_{p_{\max}}$ | $g_{p_{\min}}$ | $g_{p_{\max}}$ | $w_a$ | $\lvert\Theta^*\rvert$ | $p_{wa}$ | $f$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 5 лучших решений | | | | | | | | | | | | |
| 1 | 646 | 16 | 7765 | 4600 | 1499 | 101 | 473 | 0 | 729 | 0 | 234 | 1,023 |
| 2 | 579 | 15 | 9345 | 4940 | 1503 | 106 | 427 | 6 | 725 | 3 | 265 | 1,071 |
| 3 | 653 | 30 | 8635 | 4880 | 1466 | 123 | 527 | 0 | 742 | 0 | 249 | 1,111 |
| 4 | 577 | 23 | 9105 | 5185 | 1542 | 90 | 557 | 8 | 707 | 1 | 234 | 1,150 |
| 5 | 625 | 25 | 8695 | 5600 | 1560 | 97 | 488 | 8 | 739 | 0 | 262 | 1,156 |
| 5 худших решений | | | | | | | | | | | | |
| 6 | 620 | 57 | 8860 | 4535 | 1470 | 114 | 471 | 44 | 739 | 2 | 272 | 1,658 |
| 7 | 614 | 92 | 8735 | 4025 | 1468 | 100 | 468 | 34 | 720 | 0 | 233 | 1,716 |
| 8 | 372 | 76 | 6415 | 2950 | 1040 | 95 | 521 | 51 | 736 | 0 | 234 | 1,793 |
| 9 | 405 | 126 | 5790 | 3010 | 1046 | 112 | 552 | 61 | 726 | 0 | 213 | 2,137 |
| 10 | 407 | 150 | 5515 | 2890 | 1012 | 116 | 573 | 61 | 706 | 0 | 282 | 2,267 |

Полученные результаты прежде всего подтверждают тезис о том, что жадная стратегия имеет весьма ограниченное применение в рассматриваемой задаче построения расписания (по крайней мере, в данной реализации). Так, например, лучшее из найденных на данном этапе решений имеет 16 окон в расписании групп и несколько десятков переходов между корпусами для групп и преподавателей, имеется 473 учебных дня групп с установленной одной парой и т.д., что не позволяет использовать его на практике без существенной доработки.

С другой стороны, полученные данные показывают, что порядок рассмотрения пар при жадном построении расписания оказывает сильное влияние на качество оптимизации частных критериев качества, позволяя снижение штрафов по частным показателям в несколько раз. Потенциальным следствием этого является тот факт, что при построении большего числа решений можно надеяться на отыскание решения, более близкого к оптимуму, однако для этого потребуются существенно большие вычислительные затраты. Подобный подход также может быть использован при решении других задач из области дискретной комбинаторной оптимизации [3]. При этом его неоспоримым преимущество является хорошая распараллеливаемость в соответствии с законом Густафсона [4].

Учитывая специфику задачи и доступные вычислительные средства, для ее решения может быть эффективно использован грид. При этом входными данными для вычислительных узлов является множество пар $P$, значения $t(p_i)$, $G(p_i)$, $n(p_i)$ и начальное значение ГПСЧ $r_0$. Результирующими данными, передаваемыми на сервер проекта, являются найденные значения $a(p_i)$ и $v(p_i)$ и результаты количественной оценки качества расписания $\left(t_w, g_w, \left|\Theta^*\right|, \ldots\right)$. Объем передаваемых на сервер данных в случае необходимости может быть существенно уменьшен (с нескольких сотен килобайт до нескольких сотен байт), если возвращать лишь результаты оценки качества без возврата самого расписания. При необходимости требуемое расписание может быть построено повторно отталкиваясь от исходных данных и зная условно-оптимально значение $r_0$.

## Список литературы


1. http://en.wikipedia.org/wiki/Block_scheduling
2. Ватутин Э.И., Ватутин В.И., Романченко А.С. Оценка качества расписания вуза с использованием весовой функции // Оптико-электронные приборы и устройства в системах распознавания образов, обработки изображений и символьной информации (Распознавание – 2012). Курск: изд-во КурскГТУ, 2012. С. 136–138.
3. Ватутин Э.И. Проектирование логических мультиконтроллеров. Синтез разбиений параллельных граф-схем алгоритмов. Saarbrucken: Lambert Academic Publishing, 2011 г. 292 с. ISBN 978-3-8433-1728-3.
4. http://ru.wikipedia.org/wiki/Закон_Густавсона_—_Барсиса