***С.Н. Нестеренков**, м.т.н., начальник отдела информационных технологий ЦИИР Белорусского государственного университета информатики и радиоэлектроники*

Модель построения расписания на основе прецедентов

Введение

В статье предложен подход к построению расписания занятий вуза на основе прецедентов. Модель базируется на математическом аппарате теории графов. В основу модели положены принципы поиска и доказательства изоморфизма графов. Процесс поиска изоморфизма графов описан в терминах реляционной алгебры. Проведены экспериментальные исследования, показавшие целесообразность использования данного подхода при подготовке реального расписания, а также возможность уменьшения размерности NP-полной задачи примерно на 38,5%.Проблеме составления расписания посвящено достаточно большое количество работ [1-3]. В представленной статье ввиду трудоемкости данного процесса предлагается его разбить на несколько этапов и остановиться на одном из них.

Первый этап – это нахождение аналогий в составленных ранее расписаниях. Данные аналогии обязательно будут присутствовать, так как, как правило, учебные планы, по которым происходит обучение студентов в вузах, из года в год изменяются слабо.

Второй этап – анализ и дополнение нового расписания до вида, удовлетворяющего текущим потребностям.

В данной статье остановимся на первом этапе. Для решения данной задачи предлагается использовать математический аппарат теории графов, так как первый этап нахождения аналогий можно представить в виде процесса определения изоморфизма графов.

В настоящее время теория графов используется в таких областях, как шаблоны анализа, шаблоны распознавания образов, а также компьютерное зрение [4].

В процессе составления расписания в диспетчерскую учебного процесса вуза поступают от каждой из кафедр так называемые «сведения к расписанию», где отражен перечень дисциплин, читаемых кафедрой в текущем учебном году, с указанием того, кто из преподавателей и в каких группах будет проводить занятия.

Эти данные можно представить в виде графа. Под графом в данном случае будем понимать совокупность двух множеств вершин V и ребер $E \subset V \times V$. Множество вершин и множество ребер графа G обозначим как $V(G)$, $E(G)$ соответственно. Таким образом, так как выполняется условие $\forall \varepsilon = (a,b) \in E, (b,a) \in E$, граф будет являться неориентированным. Графом $G_1(V_1, E_1)$, где множество вершин V_1 представлено отношением R, включающим в себя занятия, проводимые в группах, а в качестве ребер E_1 – ограничения, накладываемые на одновременное проведение занятий.

Подобный граф $G_2(V_2, E_2)$ можно получить, проведя обратное преобразование уже готового составленного расписания предыдущих лет, предъявляя к нему те же требования, что и к G_1 и отбросив на данном этапе такую важную информацию как время проведения занятий и место проведения (аудиторию).

Таким образом, при данной формулировке задача сводится к поиску изоморфизма графов. Данная задача должна дать ответ на вопрос: существует ли некий подграф $G_1(V_1, E_1)$, изоморфный графу $G_2(V_2, E_2)$. То есть существуют ли множества $V \subseteq V_1$ и $E \subseteq E_1$, такие, что $|V| = |V_2|$, а также $|E| = |E_2|$, а также взаимно-однозначная функция соответствия $\phi_1 : V_2 \to V$, что $\{a,b\} \in E_2$ тогда и только тогда, когда $\{\phi_1(a), \phi_2(b)\} \in E$.

Задача изоморфизма подграфу является *NP*-полной и поэтому не имеет полиномиального решения [5].

В то же время максимальным общим подграфом (*Maximum Common Edge Subgraph* (*MCES*)) двух графов $G_1(V_1, E_1)$ и $G_2(V_2, E_2)$ называется такой граф G_{12}, который является одновременно изоморфным подграфом как для G_1, так и для G_2, содержащий при этом максимально возможное количество ребер.

Предлагается использовать следующий метод для поиска изоморфных пересечений двух графов, основанный на свойствах изоморфизма графов. Графы изоморфны тогда и только тогда, когда их матрицы смежности получаются друг из друга путем парных перестановок одинаковых строк и столбцов. Определение изоморфизма двух n-вершинных графов дает нам теоретический ответ: чтобы решить поставленную задачу, необходимо перебрать все n! взаимно однозначных соответствий между множествами вершин и установить, совмещаются ли полностью ребра графа хотя бы при одном соответствии. Таким образом, с помощью полного перебора данную задачу решить затруднительно, и требуется провести некоторые оптимизационные мероприятия.

Модель процесса построения расписания на основе прецедентов

Для формального определения постановку задачи предлагается описывать в терминах реляционной алгебры [6] и оперировать следующими множествами объектов.

Преподаватель описывается пятикомпонентным кортежем:

$$T = \{t_i\}, t_i < t_i^{flm}, t_i^r, t_i^d >, \qquad (1)$$

где $t^{flm}{}_i$ – имя, фамилия, отчество; $t^r{}_i$ – звание; $t^d{}_i$ – степень.

Группа описывается тернарным кортежем:

$$G = \{g_i\}, g_i < g_i^c, g_i^s, g_i^n > , \qquad (2)$$

где $g^c{}_i$ – курс, $g^c{}_i = \overline{1, N_c}$; $g^s{}_i$ – специальность, $g^s{}_i = \overline{1, N_s}$; $g^n{}_i$ – номер группы, $g^n{}_i = \overline{1, N_n}$. В данном случае N_c – количество курсов, N_s – количество специальностей, N_n – количество групп.

Дисциплина будет представлена бинарным кортежем:

$$S = \{s_i\}, s_i < s_i^n >, \qquad (3)$$

где $s^n{}_i$ – название дисциплины.

Тип занятий также представлен бинарным кортежем:

$$L = \{l_i\}, l_i < l_i^n >, \qquad (4)$$

где $l^n{}_i$ – название типа занятий.

Время проведения занятий представлена тернарным кортежем:

$$P = \{p_i\}, p_i < p_i^n, p_i^s, p_i^e >, \qquad (5)$$

где $p^n{}_i$ – название времени проведения занятий, $p^s{}_i$ – время начала, $p^s{}_i = \overline{1, N_s}$; $p^e{}_i$ – время окончания, pei $= \overline{1, N_e}$. В данном случае N_s – количество возможных времен начала занятия в течение учебного дня, N_e – количество возможных времен конца занятия в течение учебного дня.

День проведения занятий записывается как бинарный кортеж:

$$D = \{d_i\}, d_i < d_i^n, d_i^d >, \qquad (6)$$

где $d^n{}_i$ – название дня проведения занятий, $d^d{}_i$ – номер дня недели, $d^d{}_i = \overline{1, 7}$.

Аудитория представлена тернарным кортежем:

$$C = \{c_i\}, c_i < c_i^n, c_i^t, c_i^b >, \qquad (7)$$

где $c^n{}_i$ – название аудитории, $c^t{}_i$ – тип аудитории, $c^t{}_i = \overline{1, N_t}$; $c^b{}_i$ – корпус, в котором находится аудитория, $c^b{}_i = \overline{1, N_b}$. В данном случае N_t – количество возможных типов аудиторий, N_b – количество корпусов.

Также необходимо выделить такой агрегированный объект как «учебное занятие», включающий в себя другие объекты и описывающийся следующим образом.

Учебное занятие представляет собой семикомпонентный кортеж:

$$A = \{a_i\}, a_i < a_i^t, a_i^g, a_i^s, a_i^l, a_i^p, a_i^d, a_i^c >, \tag{8}$$

где a^t_i – преподаватель ведущий занятие, $a^t_i = \overline{1, N_t}$; a^g_i – группа в которой проводится занятие, $a^g_i = \overline{1, N_g}$; a^s_i – дисциплина, по которой проводится занятие, $a^s_i = \overline{1, N_s}$; a^l_i – тип проводимого занятия, $a^l_i = \overline{1, N_l}$; a^p_i – время проведения занятия, $a^p_i = \overline{1, N_p}$; a^d_i – день проведения занятия, $a^d_i = \overline{1, N_d}$; a^c_i – аудитория, в которой проводятся занятия, $a^c_i = \overline{1, N_c}$. В данном случае N_t – количество возможных преподавателей, N_g – количество возможных групп, N_s – количество возможных дисциплин, N_l – количество возможных типов проводимых занятий, N_p – количество возможных периодов проведения занятий, N_d – количество дней проведения занятий, N_c – количество возможных аудиторий.

Модель имеет следующие типы ограничений: типовые ограничения, пожелания, ограничения конкретного вуза. Типовые ограничения формализуем следующим образом.

1. Один и тот же преподаватель в одно и то же время может проводить не более одного занятия. Таким образом, если различные занятия проводит один преподаватель, эти занятия необходимо проводить в различные периоды времени:

$$\begin{gathered} \forall (t_i, p_j) t_i \in T, p_j \in P : (\exists \neg a_k (t_i = t_k) \wedge (a_k \in A^{p_j})) \vee \\ (\neg \exists a_k (t_i = t_k) \wedge (a_k \in A^{p_j})) \end{gathered}, \tag{9}$$

где A^{p_j} – множество занятий, проводимых во время проведения занятий p_j.

2. В одной и той же аудитории не могут проводиться одновременно несколько занятий:

$$\begin{gathered} \forall (c_i, p_j) c_i \in C, p_j \in P : (\exists \neg a_k (c_i = c_k) \wedge (a_k \in A^{p_j})) \vee \\ (\neg \exists a_k (c_i = c_k) \wedge (a_k \in A^{p_j})) \end{gathered}, \tag{10}$$

где A^{p_j} – множество занятий, проводимых во время проведения занятий p_j.

3. В любой период времени в одной и той же группе одновременно не могут проводиться различные занятия. Таким образом, если в группе необходимо провести несколько занятий, их необходимо поставить на различные периоды времени:

$$\sum a_i \leq 1, i \in A^{g_j} \cap A^{p_k} \forall (g_j, p_k) g_j \in G, p_k \in P, \qquad (11)$$

где A^{g_j} – множество занятий, проводимых в группе g_j; A^{p_j} – множество занятий, проводимых во время проведения занятий p_j.

4. Общее количество занятий в любой из периодов времени не должно быть более количества аудиторий в аудиторном фонде вуза:

$$\sum a_i \leq \sum c_j, i \in A^{c_k} j \in C, \forall p \in P, \qquad (12)$$

В качестве пожеланий могут выступать пожелания преподавателей, студентов или руководства вуза, такие как например:

1. У групп студентов не должно быть «форточек» между парами.

$$N_{P_{last}} - (N_{P_{first}} + N_P) = 0, \forall p \in P, \forall g \in G, \qquad (13)$$

где $N_{P_{last}}$ – порядковый номер последнего периода времени проведения занятий, $\overline{1, N'_p}$; $N_{P_{first}}$ – порядковый номер первого периода времени проведения занятий, $\overline{1, N'_p}$; N'_P – количество занятий на текущий день; N_p – количество возможных периодов времени проведения занятий.

2. Отсутствие учебных занятий у руководства вуза в дни, когда проводятся ректорат, научно-методический совет.

$$\forall (t_i, p_j) t_i \in T, p_j \in P \setminus P_{\text{busy}}, \qquad (14)$$

где P_{busy} – периоды времени занятости преподавателя.

В качестве ограничений конкретного вуза могут выступать следующие:

1. Не более четырех пар в день:

$$N_p \le 4, \forall g \in G, \qquad (15)$$

где N_p – количество периодов времени проведения занятий.

2. Учебная неделя шесть дней: для (6)

$$d^d_i = \overline{1,6}. \qquad (16)$$

Таким образом, решение задачи сводится к нахождению (8) на основе (1)-(7), с учетом ограничений и (9)-(16). Для решения сформулированной задачи предлагается использовать математический аппарат теории графов.

Алгоритм построения расписания занятий на основе прецедентов

Алгоритм построения расписания занятий на основе прецедентов предлагается свести к задаче определения изоморфизма графов, относящейся к классу *NP*-полных. Для решения задачи предлагается использовать следующий подход. Нам известен перечень занятий (8), и мы можем значительно упростить задачу нахождения изоморфных инвариантов. Для этого как в G_1, так и в G_2 выбираются идентичные занятия, присутствующие в обоих графах они и будут составлять перечень вершин. Далее можем произвести упорядочивание этого списка по какому-либо из критериев – например, по дисциплине и типу занятий. Упорядочив оба графа, мы можем построить матрицы смежности и найти их пересечение. Получив данное пересечение, можно вершины, вошедшие в него, брать за основу и переходить ко второму шагу составления расписания, оставленному за рамками данной статьи – дополнению расписания.

Приведем пример: Пусть два преподавателя преподают две дисциплины у двух групп студентов по следующему расписанию, представленному в таблице 1.

Таблица 1

Пример расписания занятий

№	Преподаватель	Группа	Дисциплина	Тип занятий	Время	Аудитория
1	Иванов	Группа1	Дисциплина1	Лекция	8:00-9:35	Аудитория1
2	Петров	Группа1	Дисциплина2	Лекция	9:45-11:20	Аудитория1
3	Петров	Группа1	Дисциплина2	Лекция	11:40-13:25	Аудитория3
4	Петров	Группа2	Дисциплина1	Лекция	8:00-9:35	Аудитория2
5	Иванов	Группа2	Дисциплина1	Лекция	9:45-11:20	Аудитория2

Пусть даны N множеств доменов D_1, D_2, ... D_n (№, Преподаватель, Группа, Дисциплина, Тип занятий, Время, Аудитория). Отношением R над этими множествами называется множество кортежей вида $<d_1,d_2,\ldots d_n>$, где d_1 принадлежит D_1 и т.д. Отношение R описывается таблицей 1. Составим кортежи отношения, применив следующие обозначения: № – N; Преподаватель – T; Группа – G; Дисциплина – S; Тип занятий – L; Время – T; Аудитория – C. В данном примере имеем 5 кортежей вида $K_n < N_n$, T_n, G_n, S_n, L_n, T_n, $C_n >$ отношения R.

Так как для нахождения базового расписания у нас нет необходимости в некоторых данных, применим операцию «Проекция». Таким образом, применив проекцию $\pi T_n,G_n,S_n,L_n(R)$, получаем кортежи следующего вида: $K_n < N_n$, T_n, G_n, S_n, $L_n >$ (R).

Упорядочив кортежи по некоторому перечню атрибутов, соответствующих некоторым доменам, например G_n, S_n, L_n, получим упорядоченное множество кортежей отношения R, таблица 2.

Таблица 2

Пример расписания занятий после применения проекции

№	Преподаватель	Группа	Дисциплина	Тип занятий
1	Иванов	Группа1	Дисциплина1	Лекция
2	Петров	Группа1	Дисциплина2	Лекция
3	Петров	Группа1	Дисциплина2	Лекция
4	Петров	Группа2	Дисциплина1	Лекция
5	Иванов	Группа2	Дисциплина1	Лекция

Положив, что $K'_n(R) = V_{1n}(G_1)$, изобразим полученный граф на рисунке 1.

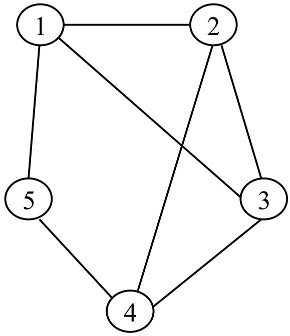

Рис. 1. Пример расписания, представленный в виде графа G_1

Для графа G_1 построим матрицу смежности вершин, таблица 3. Для простоты обозначим согласно порядковому номеру в таблице 1.

Таблица 3

Матрица смежности графа G_1

Вершины	G_1-1	G_1-2	G_1-3	G_1-4	G_1-5
G_1-1	0	1	1	0	1
G_1-2	1	0	1	1	0
G_1-3	1	1	0	1	0
G_1-4	0	1	1	0	1
G_1-5	1	0	0	1	0

Для начала работы по построению нового расписания будем использовать «сведения к расписанию», которые по структуре аналогичны данным, приведенным в таблице 2, «сведения к расписанию» приведены в таблице 4.

В данном случае также имеем 5 кортежей вида $K_n < N_n, T_n, G_n, S_n, L_n >$ отношения R'. Сразу отметим, что при предлагаемом подходе отсутствует привязка к конкретным преподавателям и отсутствуют заведомо различные кортежи отношений. Обозначим данный граф G_2, рисунок 2.

Таблица 4

Сведения к расписанию

№	Преподаватель	Группы	Дисциплина	Тип занятий
1	Преподаватель1	Группа1	Дисциплина1	Лекция
2	Преподаватель2	Группа1	Дисциплина2	Лекция
3	Преподаватель3	Группа1	Дисциплина2	Лекция
4	Преподаватель2	Группа2	Дисциплина1	Лекция
5	Преподаватель1	Группа2	Дисциплина1	Лекция

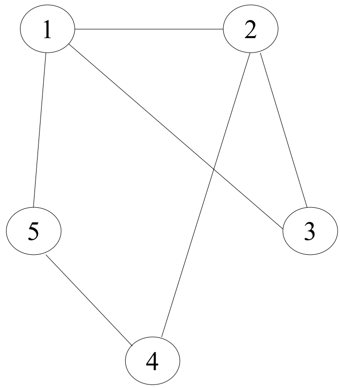

Рис. 2. «Сведения к расписанию» представленные в виде графа G_2

Метод нахождения наибольшего общего подграфа предлагается проводить по следующему алгоритму. Выбирается основной граф и граф для сравнения. Произведя обход матрицы смежности основного графа G_1, итерационно выбираем наибольшую по размеру область, идентичную области в сравниваемом графе G_2.

Таблица 5

Матрица смежности графа G2

Вершины	G_2-1	G_2-2	G_2-3	G_2-4	G_2-5
G_2-1	0	1	1	0	1
G_2-2	1	0	1	1	0
G_2-3	1	1	0	0	0
G_2-4	0	1	0	0	1
G_2-5	1	0	0	1	0

Как видим, для нашего примера пересечение графов G_1 и G_2 происходит в вершинах 1-3, (таблица 6). Таким образом, очевидно, что в качестве базового расписания необходимо взять кортежи 1-3 $K_n < N_n, T_n, G_n, S_n, L_n, T_n, C_n >$ отношения R. Проведя экспериментальные исследования на данных для первой группы одной и той же специальности одного и того же факультета только разных годов набора (2010 и 2012) для осеннего семестра, получили следующие результаты: из 14 видов занятий, проводимых в 2010, и 13, проводимых в 2012, тринадцать можно принять без изменений; при этом даже преподаватели изменились лишь в 38,5% случаев. Проведя такой же анализ для групп всего потока 2010 и 2012 годов набора, получили, что 17,85% всех занятий можно принять без изменений. Таким образом, можно сделать вывод о том, что для построения базового варианта расписания можно использовать шаблоны расписания прошлых лет и тем самым значительно сократить количество вносимой вручную информации при составлении расписания.

Таблица 6

Пересечение графов G1 и G2

Вершины	$G_{1,2}$-1	$G_{1,2}$-2	$G_{1,2}$-3	$G_{1,2}$-4	$G_{1,2}$-5
$G_{1,2}$-1	0	1	1	0	1
$G_{1,2}$-2	1	0	1	1	0
$G_{1,2}$-3	1	1	0	0	0
$G_{1,2}$-4	0	1	0	0	1
$G_{1,2}$-5	1	0	0	1	0

Как уже отмечалось выше, задача поиска изоморфизма подграфу является *NP*-полной. Для графов описывающих расписание занятий в этой задаче, удалось существенно уменьшить сложность за счет описанных выше топологических ограничений на графы, а также за счет реализованных в алгоритме модификаций, позволяющих сократить пространство поиска.

Заключение

Предложен подход к построению расписания занятий вуза на основе прецедентов. Модель базируется на математическом аппарате теории графов. В основу модели положены принципы поиска и доказательства изоморфизма графов. Процесс поиска изоморфизма графов описан в терминах реляционной алгебры. Проведены экспериментальные исследования, показавшие целесообразность использования данного подхода при подготовке реального расписания, а также возможность уменьшения размерности *NP*-полной задачи примерно на 38,5%.

Литература

1. Ерунов В.П., Морковин И.И. Формирование оптимального расписания учебных занятий в вузе. // Вестн. ОГУ. – 2001. – № 3.

2. Костенко В.А., Винокуров А.В. Локально-оптимальные алгоритмы построения расписаний, основанные на использовании сетей Хопфилда. // Программирование. – 2003. – № 4. – С.27-40.

3. Ханов Г.В., Алабужев Е.В. Автоматизация составления расписаний с учетом неопределенности.// Информационные технологии в образовании, технике и медицине. Материалы международной конференции. В 3-х т. Т.1. – ВолГТУ, Волгоград, 2004.

4. L.P. Cordella, P. Foggia, C. Sansone, M. Vento. A (Sub) Graph Isomorphism Algorithm for Matching Large Graphs IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 26, no. 10, pp. 1367-1372, 2004.

5. M.R. Garey and D.S. Johnson. Computers and Intractability: A Guide to the Theory of NP-Completeness. W. H. Freeman & Co., New York, NY, USA, 1990.

6. К.Дж. Дейт. Введение в системы баз данных. – М.: Вильямс, 2005.

S. Nesterenkov

The Model of Scheduling Based on Precedents

The paper presents an approach for university scheduling based on precedents.

The model is based on the mathematical apparatus of the theory of graphs. The model uses the principles of finding and proof of the graph isomorphism. The process of finding of the graph isomorphism described in terms of relational algebra. Experimental studies showed expediency of this approach using in the preparation of the real schedule, as well as the possibility of reducing the dimension of the NP-complete problem approximately 38.5%.

Статья поступила 23.02.2015