Рисунок 1. Реализация венгерского алгоритма в среде Borland C++ Builder

Реализация такой системы позволяет эффективно распределить сотрудников на требуемые работы. При этом план работ составлен так, что все работы выполнены, каждый сотрудник занят, а общие затраты на выполнение всех работ минимальны.

При моделировании важно понимать насколько точно модель описывает реальный объект моделирования. Для этого существуют две проверки: валидация и верификация. Их основная цель – исследование свойств имитационной модели таких как: точность, устойчивость и чувствительность результатов моделирования [1].

Таким образом, используя методы имитационного моделирования и современные информационные технологии, было получено эффективное решение задачи о назначениях.

1. Лычкина Н.Н. Имитационное моделирование экономических процессов // Учебное пособие для слушателей программы eMBI / Лычкина Н.Н., Академия АйТи, Государственный университет управления, 2005
2. Сакалаускас Л. Нелинейная стохастическая оптимизация методом Монте-Карло // Стохастическая оптимизация в информатике: Межвузовский сб. / Под ред. проф. О.Н. Граничина. – СПб.: Изд-во СПб гос. ун-та, 2005.–Вып. 1.–С.190–205.

Преображенская Т.В.

Минимизация точек ввода данных в информационных системах

*Новосибирский государственный технический университет
(Россия, Новосибирск)*

doi:10.18411/spc-12-03-2018-04

idsp: 000001:spc-12-03-2018-04

При разработке баз данных (БД) информационных систем (ИС) аналитики опираются на известные принципы [1]. Их соблюдение позволяет обеспечить: недопустимость неточно введенных данных или дублирование; защиту от ошибок при обновлении; невозможность удаления связанных данных разных таблиц; неискажение данных при работе в многопользовательском режиме и др.

Но, любая ИС в процессе ее эксплуатации претерпевает ряд изменений - растет или уменьшается количество пользователей и рабочих мест, появляются новые запросы или исчезают те, что были раньше. Изменяются интерфейсы (формы экранов и

сценарии общения с пользователем), программное обеспечение системы и др. Современные подходы к разработке программного обеспечения - Scrum и Agile, нацелены на быстрое получение результата – работающего программного продукта, разработчики практически не оставляют документации и строгих инструкций по вводу данных. Современные интерфейсы имеют тенденцию к интеллектуализации – реализации общения с пользователем на практически естественном языке. Все это приводит к тому, что утрачивается строгость контроля ввода данных (одни и те же данные по семантике могут быть введены в таблицы БД по другим синтаксическим правилам), с разных рабочих мест одни и те же данные могут быть введены по другим правилам.

Утрата жесткого контроля правил ввода данных порождает многократное дублирование потоков данных - одну из самых главных проблем современных сложных ИС. Избыточные информационные потоки увеличивают непроизводительные затраты в жизненном цикле данных (ввод, хранение, извлечение, обработка, отображение), и порождают неопределенность, хаос, недоверие к отчетам, трудности интерпретации неоднозначных данных и другие проблемы, оказывающие влияние на качество данных. В результате база данных засоряется - накапливаются дублированные хранимые данные.

Один из подходов решения проблемы - использование современных средств очистки данных, используемых в технологиях добычи данных (Datamining). Возникает вопрос – зачем создавать неопределенность в структурированных данных, чтобы потом ее снимать с помощью интеллектуальных систем? Не проще ли иногда наводить порядок в данных. Один из приемов – использование систем управления взаимоотношениями с клиентами (CRM), когда в единой точке собираются корректные данные о потребителях, вводятся в БД и становятся централизованно доступными всем пользователям системы.

Предлагается еще один прием – предварительная семантическая проверка вводимых данных, для использования в системе сбалансированных показателей при информационном аудите системы. ИС периодически подвергается аудиту – необходимой процедуре для выявления соответствия ИС текущим информационным потребностям бизнеса, и далее - реинжинирингу (перестройке процессов сбора и алгоритмов обработки данных, изменениям концептуальной схемы БД и др.). Предварительная семантическая обработка данных на этапе аудита позволит выявить минимально необходимое количество вводимых данных, а далее и точек их ввода, что приведет, в конечном счете, к значительному снижению непроизводительных затрат и неопределенности. Глубокая аналитическая работа с данными - залог успеха проекта в целом (создание долгоживущей концептуальной схемы БД, которая в процессе эксплуатации не перестраивается, а развивается и наращивается).

Минимизация окончечных точек для ввода данных позволит значительно сократить дублирование данных при их вводе и хранении [2]. В реальных организациях множество ключевых показателей может включать в себя до сотни и больше отдельных показателей, записанных в разных обозначениях (примеры их записи даны в 4-м столбце таблицы). Необходима большая работа по анализу семантики показателей и их операндов, приведению их к единым обозначениям (примеры в столбце 5 таблицы). Свести данные в единую интегрированную систему, готовую к алгоритмизации и хранению в базе данных – задача аналитиков. Появление всякого нового показателя потребует вновь экспертизы семантической новизны самого показателя, атрибутов, операций и др. Деятельность аналитиков по выявлению точек ввода в ИС можно частично автоматизировать на основе предложенных алгоритмов.

Если представить каждый отдельный вычисляемый показатель иерархическим графом, в котором главная вершина – конечный результат вычислений, а другие вершины – операнды формулы, то на основе объединения элементарных (частных) графов можно получить избыточную сеть вычисления показателей организации. Например, для показателя $\Pi = A1 + A2$, а для показателя $A2 = A21 + A22$ можно

записать общую формулу вычисления показателя Π в виде орграфа $\Pi(A1, A2(A21, A22))$ [2,3,4]. Правило чтения записи орграфа – строго сверху вниз и слева направо на каждом уровне иерархии; скобка – начало нового уровня, запятая – разделитель при перечислении вершин на одном уровне, число пар скобок равно уменьшенному на единицу числу уровней иерархии. Алгоритм формирования обобщенного графа состоит из следующих шагов:

Шаг 1. Для каждого отдельного показателя строится частный орграф, все вершины графа получают обозначение в соответствии с принятой единой системой обозначений (последний столбец таблицы).

Шаг 2. Каждый полученный иерархический граф записывается аналитически в виде матрицы $A = \|a_{ij}\|$, где $a_{ij} = 1$, если при вычислении вершины i используется вершина j и $a_{ij} = 0$, если при вычислении вершины i не используется вершина j . Матрица квадратная.

Шаг 3. Строится обобщенная матрица на основе включения всех строк и столбцов частных матриц без повторения (путем наложения частных графов).

Полученная обобщенная матрица – аналитическая запись орграфа системы показателей и ее можно обработать на вычислительной машине. Проиллюстрировать работу алгоритма можно на простом примере (таблица). Пусть мы имеем несколько простых показателей, представленных частными графами: $\Pi_1(\Pi_3, A_1)$; $\Pi_2(\Pi_3, A_2)$; $\Pi_3(A_3, A_4)$; $\Pi_4(\Pi_3, A_5(A_{11}, A_{15}, A_{16}))$; Π_5 ; $\Pi_6(A_6, A_7)$; $\Pi_7(A_3, A_8)$; $\Pi_8(A_9, A_{10})$; $\Pi_9(A_{11}, A_{12})$; $\Pi_{10}(\Pi_9(A_{11}, A_{12}), A_{13})$; $\Pi_{11}(A_{14}, A_{12})$; Π_{12} ; $\Pi_{13}(\Pi_{12}, A_{12})$. Выделение минимального набора хранимых данных (МНХД) позволит устранить дублирование данных, сократить объем хранимых в БД данных, необходимых для расчета показателей. В рассмотренном исходном варианте количество всех показателей и операндов составляло 35 (таблица). Выявление МНХД позволило сократить количество данных, необходимых для расчета, а, следовательно, и хранения, до 17 исходных значений. Приведенный алгоритм обеспечит согласованность данных в едином информационном пространстве вычислений.

Таблица

Фрагмент системы показателей организации и новые обозначения для интегрированной системы показателей

Группа системы показателей	Показатель	Единица измерения	Расчетные формулы в обозначениях разных специалистов	Соответствие обозначений специалистов новым обозначениям операндов в интегрированной системе показателей
1	2	3	4	5
Финансы	Капиталоотдача	%	$KO = \frac{B\P}{K} * 100\%$	КО – Π_2 ; К – A1
	Оборачиваемость текущих активов	%	$OTA = \frac{B\P}{TA} * 100\%$	ОТА – Π_2 ; ТА – A2
	Выручка от продаж	тыс. руб.	$B\P = V_{\Sigma} * P$	$B\P$ – Π_3 ; V_{Σ} – A3; P – A4
	Валовая прибыль	тыс. руб.	$\Pi_4 = B\P - C / C (C = V * C_{\Sigma} + C_{\Sigma} \text{ст})$	Π_4 – Π_4 ; C/C – A5; C_{Σ} – A15; $C_{\Sigma} \text{ст}$ – A16
Клиенты	Количество постоянных потребителей	тыс. чел.	Q	Q – Π_5
	Индекс удовлетворенности потребителей	%	$I = \sum_{j=1}^n \varphi(x_j) q_j$	I – Π_6 ; $\varphi(x_j)$ – A6; q_j – A7
	Прирост объема продаж	%	$\Delta V = \frac{V_{\Sigma} - V_P}{V_P} * 100\%$	ΔV – Π_7 ; V_P – A8
Внутренние бизнес-процессы	Длительность производственного цикла	день	$T_{\Sigma} = T_{\Sigma} \text{ц} + T_{\Sigma} \text{о}$	T_{Σ} – Π_8 ; $T_{\Sigma} \text{ц}$ – A9; $T_{\Sigma} \text{о}$ – A10
	Производительность труда	шт./мес.	$Pa = \frac{V}{N_z}$	Pa – Π_9 ; V – A11; N_z – A12
	Рост производительности труда	%	$\Delta Pa = \frac{Pa - Pa_2}{Pa_2} * 100\%$	ΔPa – Π_{10} ; Pa_2 – A13
Обучение и развитие	Коэффициент оборота по приему	-	$K_{\text{оп}} = \frac{N_1}{N_z}$	$K_{\text{оп}}$ – Π_{11} ; N_1 – A14
	Расходы на обучение	тыс. руб.	$C_{\text{об}}$	$C_{\text{об}}$ – Π_{12}
	Затраты на обучение 1 работника	тыс. руб.	$C_{\text{об,к}} = \frac{C_{\text{об}}}{N_z}$	$C_{\text{об,к}}$ – Π_{13}
<p>Толкование обозначений в алгоритмах: B\P – выручка от продаж, K – вложенный капитал, TA – текущие активы, V_{Σ} – объем продаж в текущем периоде, P – цена единицы продукции, Π_4 – валовая прибыль, C/C – себестоимость, $\varphi(x_j)$ – относительное значение соответствия j-го показателя качества x, q_j – экспертный весовой коэффициент расчета комплексных показателей для j-го показателя множества $\{\varphi\}$, V_P – объем продаж в предыдущем периоде, $T_{\Sigma} \text{ц}$ – длительность операционного цикла, $T_{\Sigma} \text{о}$ – длительность межоперационных перерывов, V – выпуск продукции, N_z – среднесписочная численность работников, N_1 – количество принятого персонала.</p>				

Работа алгоритмов, написанных на VBA, опробована в среде MS Excel. Обобщенную матрицу можно также использовать для автоматизированного построения вычислительных процедур при расчете показателей. Предложенный подход позволит значительно сократить затраты времени, уменьшить количество ошибок при анализе данных, выявить МНХД для хранения и реализации всех расчетов. Кроме того, приведенные алгоритмы очень полезны при обучении студентов основам проектирования баз данных для иллюстрации идеи интегрирования данных в единую систему.

1. Бойко В.В., Савинков В.М. Проектирование баз данных информационных систем. - 2-е изд., перераб. и доп. - М.: Финансы и статистика, 1989.-351с.
2. Преображенская Т.В. Формализация деятельности аналитиков, работающих в сфере бизнеса //Автомобильная промышленность, 2010, № 7, с.5-8
3. Каплан Р. С., Нортон Д. П. Сбалансированная система показателей. От стратегии к действию. / Изд. 2-е, испр. и доп.: Пер. с англ. - М.: ЗАО «Олимп-Бизнес», 2003.- 320 с.
4. Денисов А. А. Теория больших систем управления : учебное пособие для вузов по специальности "Автоматика и телемеханика"/ А.А. Денисов, Д.Н. Колесников. Л.: Энергоиздат , 1982.- 287 с.