УДК 618.513

А.С. Лихоузов, Т.А. Лихоузова

ИСПОЛЬЗОВАНИЕ МЕТОДОВ ТЕОРИИ ИДЕНТИФИКАЦИИ ДЛЯ АНАЛИЗА И ОПТИМИЗАЦИИ ВЫЧИСЛИТЕЛЬНЫХ ПРОЦЕССОВ

Введение

Методы теории идентификации очень полезны для повышения производительности приложений основывающихся на применении языка SQL (PL/SQL). С их помощью можно повысить эффективность инструментальных средств диагностики, входящих в Oracle, и выявить скрытые проблемы, которые нельзя увидеть с помощью других методов, а также спрогнозировать производительность приложения при более высоких нагрузках.

Анализ проблемы

Рассмотрим широко распространенную методику анализа проблемного SQL-кода по модели производительности, которая предложена в [1].

Шаг1. Изолирование рассматриваемого кода SQL.

Рассматриваемый SQL-код изолируется от окружающего системного кода и помещается в SQL*PLUS или сценарий PL/SQL, независимое выполнение которого позволит имитировать производственный процесс.

Шаг 2. Тестирование в идеальных условиях.

Идеальные условия означают один процесс SQL, выполняющийся на выделенной машине с аппаратными средствами, имеющими фиксированную мощность обработки данных и работающими с большим объемом данных.

Шаг 3. Построение графика по наблюдаемым значениям производительности относительно осей координат строка-время.

Форма тренда может подсказать причину основных проблем производительности.

Шаг 4. Использование простого определения уравнения.

После нанесения точек на график можно предположить, что прямая линия — это линейная функция, а направленная вверх кривая — квадратическая (могут быть и другие формы, но они, как правило, не рассматриваются). Исходя из этих наблюдений, для определения уравнений можно использовать методы [2], основанные на упрощенных вариантах интерполяционного многочлена бинома Ньютона (либо простой двухточечный линейный метод, либо трехточечный квадратический метод).

Шаг 5. Прогнозирование производительности.

Полученные уравнения позволяют прогнозировать производительность при намного более высоких нагрузках, чем те, что тестируются на практике. Поскольку точность прогнозирования существенно уменьшается по мере увеличения нагрузки, этот метод предназначен только для приблизительных прогнозов.

Шаг 6. Интерпретация модели производительности и проведение эксперимента.

Форма линий производительности и характер уравнений указывают путь к решению основных проблем производительности [3].

Шаг 7. Запись результатов для проведения экспертизы.

При использовании этой методики далеко не всегда есть возможность проводить тестирование в идеальных условиях. Производительность компьютера может быть разной: в течение обычного рабочего времени возникают разные ситуации с рабочей нагрузкой; необходимо также учитывать, что число записей в базе данных может возрасти за последующие месяцы/годы. Может возникнуть ситуация, при которой запрос выполняется, когда процессоры и ОЗУ заняты другими приложениями или когда многие приложения (выполнение которых не предполагалось заранее) могут внезапно привлекать ресурсы. Это приводит к тому, что на четвертом шаге при использовании методов [2] будет построена модель системы, в которой оценки параметров существенно отличаются от истинных значений. Следовательно, ни о каком точном прогнозе на пятом шаге не может быть речи.

Постановка задачи

Для корректного использования приведенной выше методики необходимо применение более точных моделей процесса. Решением в данной ситуации может стать использование более сложных методов для построения модели– например, методов теории идентификации. Эти методы рассчитаны на работу с зашумленными данными, поэтому дадут более точный результат. Ограничением для их применения к решению данной задачи является невозможность проводить большое количество замеров времени выполнения кода на больших объемах данных.

Среди алгоритмов идентификации наиболее подходящими для решения данной задачи являются рекуррентный регрессионный метод, различные варианты метода стохастических аппроксимаций и метод последовательного обучения [4, 5, 6]. При этом простота и универсальность последних позволяет не ограничиваться квадратичными критериями идентификации и формировать различные как линейные, так и нелинейные алгоритмы идентификации.

Регрессионная идентификация. Алгоритмы идентификации, основанные на регрессионных процедурах с использованием метода наименьших квадратов, применимы как к линейным, так и нелинейным процессам и системам. Преимущество рекуррентных регрессионных алгори-

 ISSN 1562-9945

тмов идентификации по сравнению с классическими в том, что при наращивании количества наблюдений никакие предшествующие данные о процессе не игнорируются, а предыдущие оценки параметров корректируются, выступая в рекуррентном алгоритме в качестве начального приближения.

Математическая модель линейной системы с m-входами (одновременно обрабатываемыми таблицами) и одним выходом (временем обработки) может быть представлена следующим уравнением:

$$x = \sum_{j=0}^{m} a_j \cdot u_j, \tag{1}$$

где x – выход системы (время выполнения), u_j – входное воздействие (количество строк в j-той таблице), причем $u_0 = 1$, a_j – параметры системы.

Так как на практике, как правило, переменная x оказывается недоступной для измерения (или наблюдения), то при идентификации вместо x используется связанная с ней некоторая переменная $z = x + n$, где n – случайная величина с нулевым средним и конечной дисперсией.

На интервале наблюдения фиксируется совокупность дискретных значений наблюдаемых величин входа и выхода u_{ij}, $z_i = x_i + n_i$ ($j = 0$, m, i=1, N). Вся зафиксированная совокупность из N измерений входа-выхода представляется в виде следующего векторно-матричного уравнения:

$$\bar{z} = U\bar{a} + \bar{n}, \tag{2}$$

где $\bar{z}^T = [z_1, z_2, \ldots, z_N]$ — N-мерный вектор наблюдаемых значений выходных сигналов, $\bar{n}^T = [n_1, n_2, \ldots, n_N]$ — N-мерный вектор помех, $\bar{a}^T = [a_0, a_1, \ldots, a_m]$ — m-мерный вектор параметров системы, U - матрица $N \times (m+1)$ входных воздействий следующего вида:

$$U = \begin{bmatrix} u_{01} & u_{11} & \ldots & u_{m1} \\ u_{02} & u_{12} & \ldots & u_{m2} \\ \ldots & \ldots & \ldots & \ldots \\ u_{0N} & u_{1N} & \ldots & u_{mN} \end{bmatrix} = \begin{bmatrix} \bar{u}_1^T \\ \bar{u}_2^T \\ \ldots \\ \bar{u}_N^T \end{bmatrix} \tag{3}$$

Обозначив как $\hat{\bar{a}}$ вектор оценок искомых параметров системы, а как $\hat{\bar{x}}$ — вектор оценок значений выходной переменной системы; идентифицирующая модель представляется как

$$\hat{\bar{x}} = U\,\hat{\bar{a}} \tag{4}$$

Мерой близости идентифицирующей модели и системы является квадратичная форма, представляющая собой сумму квадратов разности между значениями выхода системы и идентифицирующей модели, т.е.

$$I = \sum_{i=1}^{N} (z_i - \hat{x}_i)^2 \tag{5}$$

Значения вектора оценок параметров системы находятся из условия минимума критерия идентификации (5), а именно:

$$\frac{\partial I}{\partial \hat{\bar{a}}} = \frac{\partial \left((\bar{z} - U \cdot \hat{\bar{a}})^T \cdot (\bar{z} - U \cdot \hat{\bar{a}}) \right)}{\partial \hat{\bar{a}}} = 0, \tag{6}$$

откуда

$$U^T \cdot U \cdot \hat{\bar{a}} = U \cdot \bar{z} \tag{7}$$

Решение системы уравнений (7) дает оптимальные значения оценок параметров:

$$\hat{\bar{a}}^* = (U^T \cdot U)^{-1} \cdot U \cdot \bar{z}, \tag{8}$$

что будет иметь место при условии, если матрица U неособенная.

При рекуррентном регрессионном оценивании параметров в критерий идентификации часто вводят весовые коэффициенты q_i, $0 \leqslant q_i \leqslant 1$, учитывающие значимость различных переменных в исходном наборе (зачастую их можно задать, проанализировав загрузку задействованных аппаратных средств во время замера), то есть функционал (5) принимает следующий вид:

$$I_N = \sum_{i=1}^{N} q_i (z_i - \hat{x}_i)^2 \tag{9}$$

Обычно при равнозначности наблюдений принимается, что для каждого $(i = \overline{1, N})$, $q_i = 1$, а рост значений q_i при $i \to N$ увеличивает вес последних наблюдений. Здесь N указывает на объем выборки, при котором решается задача оценивания, а само значение N может последовательно увеличиваться.

Используя, как и ранее, необходимое условие экстремума (минимума) функционала, представляем решение системы линейных алгебраических уравнений (7) в виде следующих рекуррентных соотношений:

$$\hat{\bar{a}}_k = \hat{\bar{a}}_{k-1} + p_k \cdot q_k \cdot \bar{u}_k \cdot (z_k - \bar{u}_k^T \cdot \hat{\bar{a}}_{k-1}) \quad (k = \overline{1, N}), \tag{10}$$

где

$$p_k^{-1} = \sum_{i=1}^{k} q_i \cdot (\bar{u}_i \cdot \bar{u}_i^T) = p_{k-1}^{-1} + q_k \cdot \bar{u}_k \cdot \bar{u}_k^T; \forall k \geqslant 1 \tag{11}$$

Выражения (10) и (11) необходимо дополнить начальными значениями вектора оценок $\hat{\bar{a}}_0$ и матрицы p_0, которые могут быть приняты следующими:

$$\hat{\bar{a}}_0 = 0;\ p_0 = \frac{1}{\varepsilon} \cdot I, \text{ при } \varepsilon \to 0 \tag{12}$$

Значение множителя $\frac{1}{\varepsilon}$ может быть выбрано произвольным, в диапазоне от 10 до максимально возможного.

Идентификация методом стохастической аппроксимации. Математическая модель системы определяется уравнением вида:

$$x = f(\bar{a}, \bar{u}) \tag{13}$$

где x – переменная выхода; $\bar{u}^T = [u_1, \ldots, u_m]$ – вектор переменных m-входов; $\bar{a}^T = [a_1, \ldots, a_m]$ —m-мерный вектор неизвестных параметров системы; $f(\bar{a}, \bar{u})$ — непрерывная (в общем случае нелинейная функция своих переменных). В частном случае, когда система линейна, уравнение (13) может быть представлено в виде (1). Модель наблюдения: $z = x + n$.

Полагая, что наблюдения значений входа-выхода системы производится в установившемся режиме в фиксированные моменты времени $t_k \in [0, T], k = \overline{0, N}$, приходим к дискретной форме представления уравнения системы:

$$\begin{cases} x_k = f(\bar{a}, \bar{u}_k) \\ z_k = x_k + n_k \end{cases} \quad (k = \overline{0, N}); \tag{14}$$

Тогда оцениваемая модель системы:

$$\hat{x}_k = f(\hat{\bar{a}}, \bar{u}_k), k = \overline{0, N}, \tag{15}$$

где $\hat{\bar{a}}^T = [\hat{a}_1, \ldots, \hat{a}_m]$ – вектор оценок параметров системы $a_j, j = \overline{1, m}$.

Вводим критерий идентификации следующего вида:

$$I_k(\hat{\bar{a}}) = 0,5[z_k - f(\hat{\bar{a}}_k, \bar{u}_k)]^2 \tag{16}$$

Тогда, в соответствии с алгоритмом стохастической аппроксимации (алгоритм Кифера-Вольфовица [4]) оценка $\hat{\bar{a}}_k$ вектора $\bar{a}$ (т.е. на k-ом последовательном шаге) определяется следующим соотношением:

$$\hat{\bar{a}}_{k+1} = \hat{\bar{a}}_k - \rho_k \cdot \bar{\psi}_k; \quad \forall k = \overline{1, N}; \tag{17}$$

где

$$\bar{\psi}_k^T = \left[\frac{\partial I_k(\hat{\bar{a}})}{\partial \hat{\bar{a}}_k}\right]^T = \left[\frac{\partial I_k(\hat{\bar{a}})}{\partial \hat{a}_{1k}}, \frac{\partial I_k(\hat{\bar{a}})}{\partial \hat{a}_{2k}}, \ldots, \frac{\partial I_k(\hat{\bar{a}})}{\partial \hat{a}_{mk}}\right] \tag{18}$$

Учитывая выражение (16), представляем (18) следующим образом:

$$\bar{\psi}_k = \frac{\partial I_k(\hat{\bar{a}})}{\partial \hat{\bar{a}}_k} = \left[f(\bar{u}_k, \hat{\bar{a}}_k) - z_k\right] \cdot \frac{\partial f(\bar{u}_k, \hat{\bar{a}}_k)}{\partial \hat{\bar{a}}_k} = \left[f(\bar{u}_k, \hat{\bar{a}}_k) - z_k\right] \cdot g_k(\bar{u}_k, \hat{\bar{a}}_k), \tag{19}$$

где $g_k(\bar{u}_k, \hat{\bar{a}}_k) = \frac{\partial f(\bar{u}_k, \hat{\bar{a}}_k)}{\partial \hat{\bar{a}}_k}$;

Итерационный процесс оценивания, определяемый (17), требует задания начального значения оценки $\hat{\bar{a}}_0$ и последовательности коэффициентов ρ_k, которую определяем как:

$$\rho_k = \frac{\rho_0}{1 + k}; (k = 1, 2, \ldots, N) \tag{20}$$

В общем, нелинейном, случае нет рекомендаций по выбору начальных значений $\hat{\bar{a}}_0$, ρ_k, но для линейной системы можно принять следующие значения:

$$\hat{\bar{a}}_0 = 0; \rho_0 = \left[g^T(\hat{\bar{a}}_0, \bar{u}_0) \cdot g(\hat{\bar{a}}_0, \bar{u}_0) \right]^{-1} \quad (21)$$

Процедура оценивания, определяемая соотношениями (17)-(19), существенно упрощается для линейной системы, так как при $x_k = \bar{a}^T \cdot \bar{u}_k$ вектор $\bar{\psi}_k$ будет подчиняться уравнению:

$$\bar{\psi}_k = (\hat{\bar{a}}^T \bar{u}_k - z_k) \cdot g(\bar{u}_k, \hat{\bar{a}}_k) = (\hat{\bar{a}}^T \bar{u}_k - z_k) \cdot \bar{u}_k \quad (22)$$

Идентификация методом последовательного обучения. Данный метод основан на последовательных (пошаговых) процедурах оценивания параметров модели, аналогичен методу стохастической аппроксимации и отличается от него большей сходимостью, простотой алгоритма и может быть применен к линейным системам с медленно меняющимися параметрами. В силу этого, оценка, полученная методом последовательного обучения, может быть использована в качестве начальной оценки для других последовательных процедур идентификации, таких как метод стохастической аппроксимации.

Рассмотрим применение данного метода для оценивания параметров линейной стационарной системы с одним выходом и несколькими входами. Математическая модель системы - уравнение вида (1).

Полагая, что производится ряд последовательных измерений совокупности вход/выход, имеем:

$$\begin{cases} x_k = \bar{a}^T \cdot \bar{u}_k; \\ z_k = x_k + n_k \end{cases} ; \ (k = 0, 1, \ldots, N) \quad (23)$$

Оцениваемая модель системы имеет вид:

$$\hat{x}_k = \hat{\bar{a}}^T \cdot \bar{u}_k \quad (k = 0, 1, \ldots N), \quad (24)$$

а алгоритм оценивания параметров определяется следующим образом:

$$\hat{\bar{a}}_{k+1} = \hat{\bar{a}}_k + \frac{(x_k - \hat{x}_k)}{\|\bar{u}_k\|^2} . \bar{u}_k \ (k = 0, 1, \ldots, N), \quad (25)$$

где $\|\bar{u}_k\|$ – эвклидова норма вектора, $\|\bar{u}_k\|^2 = \bar{u}_k^T \cdot \bar{u}_k$.

При отсутствии какой-либо априорной информации о начальных значениях оцениваемых параметров, их можно принять нулевыми $\hat{\bar{a}}_0 = 0$, что соответственно обуславливает также нулевые значения выхода $\hat{x}_0 = 0$.

В уравнении (25) допустимо введение коэффициента коррекции ошибки идентификации α (дополнительного коэффициента пропорциональности) следующим образом:

$$\hat{\bar{a}}_{k+1} = \hat{\bar{a}}_k + \alpha \cdot \frac{(x_k - \hat{x}_k)}{\bar{u}_k^T \cdot \bar{u}_k} \cdot \bar{u}_k \quad (k = 0, 1, \ldots, N), \quad (26)$$

где $0 < \alpha < 2$; что гарантирует сохранение сходимости.

Алгоритм (25) получен, как и ранее, в методе стохастической аппроксимации, из условия минимизации на каждом шаге квадрата мгновенной

ошибки оценивания $I_k^{=}0,5 \cdot \left\|\hat{\bar{a}}_k - \bar{a}\right\|^2$, когда в качестве коэффициента пропорциональности ρ_k принимается:

$$\rho_k = \frac{1}{\left\|\bar{u}_k\right\|^2}; \quad k = \overline{1, N} \tag{27}$$

В качестве условия окончания итерационного процесса идентификации может быть принято следующее:

$$\left\|\Delta\hat{\bar{a}}_{k+1}\right\|^2 = \left\|\hat{\bar{a}}_{k+1} - \hat{\bar{a}}_k\right\| \leqslant \varepsilon, \tag{28}$$

где ε – заданная погрешность идентификации параметров.

Интерпретация моделей. Графические модели производительности дают ключ к устранению основных проблем SQL-кода (рис. 1).

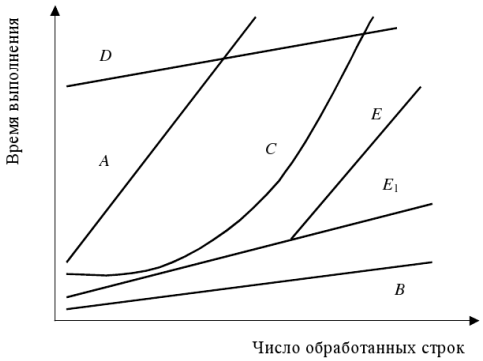

Рис. 1 – Графические представления различных ситуаций при оптимизации

Конечная цель — с помощью оптимизации кода преобразовать крутую линейную или квадратичную линию оптимальной производительности в пологую и линейную. Для этого могут потребоваться эксперименты с индексами, временными таблицами, командами с инструкциями оптимизатору или другими методами оптимизации производительности SQL. В таблице 1 представлены советы, которые дают общую идею относительно того, как можно применить наблюдаемые зависимости непосредственно к оптимизации кода SQL [3].

Выводы

Объединение методов теории идентификации с диагностическими инструментальными средствами Oracle позволяет четче интерпретировать модели производительности и повышать эффективность кода, поскольку помогает идентифицировать скрытые проблемы, которые другие диагностические методы могут пропустить. Кроме того, такой подход

Таблица 1.

Модель	Возможная проблема	Возможное решение
A	Отсутствие индекса для запроса, выбирающего значения	Создайте индекс. Восстановите подавленный индекс.
A	Таблица, у которой слишком много индексов, неэффективна во время выполнения команд DML.	Удалите некоторые из индексов, или индексируйте меньшее число столбцов (или столбцы меньшего размера) для текущих индексов.
B	Нет проблем.	-
C	Отсутствие индекса для запроса, выбирающего значения.	Создайте индекс. Восстановите подавленный индекс.
C	Таблица, у которой слишком много индексов, неэффективна во время выполнения команды INSERT.	Удалите некоторые из индексов или индексируйте меньшее число столбцов (или столбцы меньшего размера) для текущих индексов.
D	Выполнение полного просмотра таблицы или использование инструкции ALL_ROWS, когда этого не следует делать.	Попробуйте выполнить поиск по индексу. Попытайтесь использовать инструкцию FIRST_ROWS для принудительного использования индексов.
E	Запрос отлично работал, пока не столкнулся с другим ограничением (например, дисковый ввод/вывод или в памяти).	Выясните, какое максимальное достигаемое значение вызывает эту проблему. Эту проблему может решить увеличение SGA, но она может быть вызвана многими другими факторами.
E_1	Если ограничение на линии *E* исправлено, обработка должна продолжиться по прямой линии.	Дальнейшая оптимизация может улучшить процесс до линии *B*.

 ISSN 1562-9945

помогает преодолеть барьер оптимизации производительности, связанный с неопытностью в работе с Oracle, недостатком явных фактов или трудностями с интерпретацией диагностического инструмента.

Литература

1. Holmes, J.A., "Seven Deadly SQL Traps and How to Avoid Them"/ SELECT Magazine, Vol. 6, 4, July 1999, IOUG-A, pp. 22-26.
2. Holmes, J.A., "Leveraging Oracle Performance Tuning Tools Using Simple Mathematical Techniques"/ SELECT Magazine, Vol. 5, 4, July 1998, IOUG-A, pp. 36-42.
3. Ниемек Р.Дж. "Oracle9i. Оптимизация производительности. Советы и методы"/ М.: Издательство "Лори" Издательский дом "Питер", 2007-726с.
4. Гроп Д. "Методы идентификации систем"/ М., Мир, 1979.
5. Сейдж А., Мелса Дж "Идентификация систем"/ М.,Наука,1974.
6. Эйкхофф П., "Основы идентификации систем управления"/ М., Мир, 1975.

Получено 03.04.2008