

## МЕТОД ОПТИМИЗАЦИИ РАСПРЕДЕЛЕННЫХ БАЗ ДАННЫХ

Лаздынь С.В., Телятников А.О.

Донецкий национальный технический университет,  
кафедра автоматизированных систем управления  
E-mail: Alexander.Telyatnikov@gmail.com

### Abstract

*Lazdyn S.V. Telyatnikov A.O. Method of the distributed databases optimization . In the article the method of optimization of Distributed Data Bases (DDB) is offered based on sharing of object model of theDDB and modified Genetic Algorithm (GA). The scheme of co-operation of the DDB model with GA is described. Modification of genetic algorithm principle of encoding of decisions and genetic operators are considered. The algorithm of optimization of distributed databases on the nodes of network is described.*

### Введение

Создание компьютерных информационных систем крупных предприятий все чаще предполагает использование распределенных баз данных (РБД), в которых данные распределены по множеству узлов при помощи фрагментации и репликации. Производительность РБД зависит как от параметров технических средств (серверов и каналов передачи данных), так и от того насколько оптимально распределены данные в системе.

Вопросам моделирования и оптимизации РБД посвящен ряд научных работ Г.Г. Цегелика [1], А.Г. Мамиконова, В.В. Кульбы [2] и других ученых. Предложенные в этих работах методы оптимизации затруднительно использовать для реальных сложных РБД с большим числом узлов обработки и таблиц БД, так как они требуют значительных затрат времени и вычислительных ресурсов.

Таким образом, несмотря на проведенные исследования задача разработки методов оптимизации РБД не получила окончательного решения и является актуальной.

### Оптимизация РБД

Для оптимизации распределения данных по узлам сети предложен новый подход, основанный на совместном использовании объектной модели РБД и аппарата ГА [3, 4]. Общая объектная модель РБД построена как система взаимодействующих объектов ее типовых компонентов: узел, канал передачи данных, приложение, запрос, таблица [3].

В качестве критерия эффективности РБД предложено использовать суммарное среднее время выполнения запросов и распространения обновлений, порожденных функционированием системы в течении заданного интервала времени, которое определяется следующим выражением:

$$T = \frac{1}{N_q} \sum_{s=1}^{N_q} t'_s + \frac{1}{N_u} \sum_{e=1}^{N_u} t''_e. \quad (1)$$

где  $N_q$  — количество запросов в системе;  $N_u$  — количество обновлений в системе;  $t'_s$  — время выполнения  $s$ -го запроса,  $s \in [1, N_q]$ ;  $t''_e$  — время распространения  $e$ -го обновления данных,  $e \in [1, N_u]$ .

Учитывая недостатки ранее применявшихся классических методов, для решения сформулированной задачи оптимизации РБД предложено использовать генетические алгоритмы.

Схема распределения фрагментов данных по узлам РБД кодируется в виде хромосомы ГА. Популяция ГА представляет собой набор некоторых точек пространства поиска.

Начальная популяция генерируется случайным образом. В процессе оптимизации с помощью операторов ГА генерируются хромосомы, то есть схемы распределения данных. Полученные схемы являются исходной информацией для объектной модели, с помощью которой вычисляются оценки критерия эффективности РБД. Эти оценки, в свою очередь, являются значениями функции приспособленности ГА (целевой функции) для данного варианта решения. То есть, другими словами, предлагается использовать разработанную объектную модель РБД для вычисления функции приспособленности ГА. Схема взаимодействия модели РБД с ГА представлена на рис. 1.

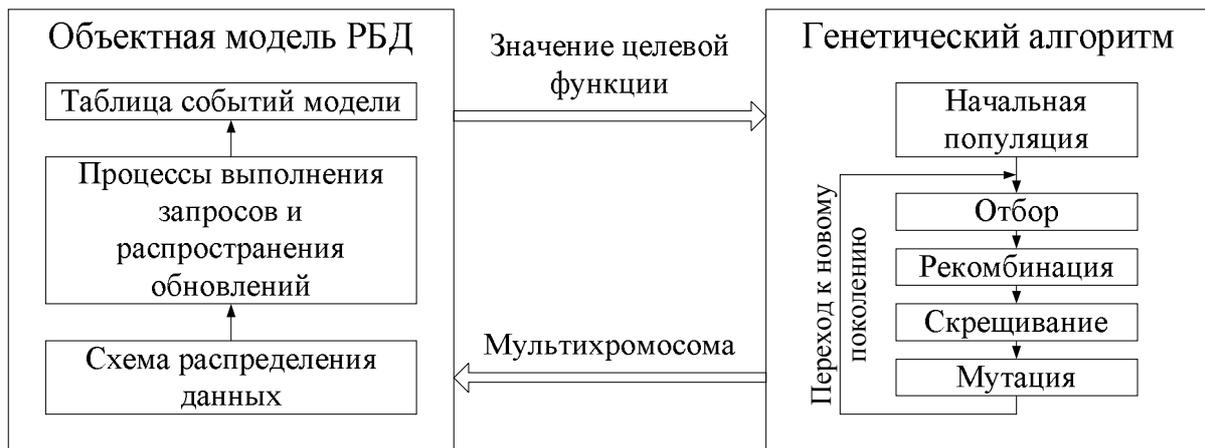


Рисунок 1 — Схема взаимодействия модели РБД с ГА

### Модификация генетического алгоритма

При разработке ГА для поиска субоптимальных решений поставленной задачи необходимо учитывать специфику этой задачи. Прежде всего, это касается структуры хромосомы, принципов ее кодирования и декодирования. Для данной задачи предложено использовать мультихромосомное представление решений [5]. Количество хромосом, необходимых для кодирования схемы распределения данных по узлам РБД будет равным числу узлов.

Каждая хромосома представляет собой вектор бинарных значений, длина которого равна количеству фрагментов данных подлежащих распределению. Элементы вектора принимают значение “1”, если копия соответствующего фрагмента данных находится на узле, и “0” в противном случае. Таким образом, каждая хромосома отвечает за определенный аспект решения, а именно, она кодирует размещение фрагментов данных на отдельно взятом узле. Весь набор хромосом полностью кодирует схему размещения данных в РБД (рис. 2).

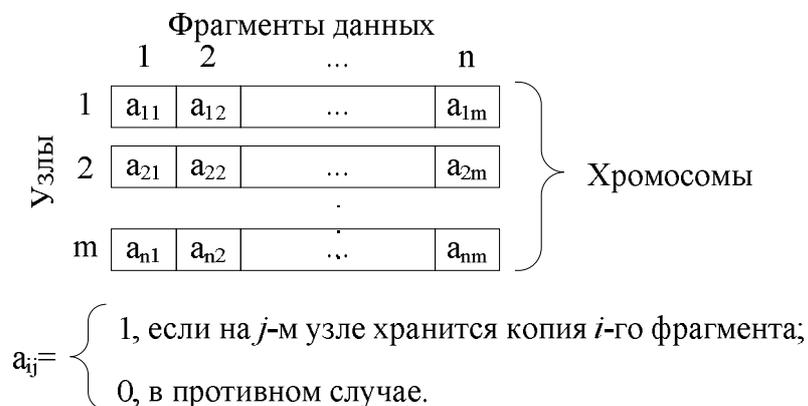


Рисунок 2 — Структура мультихромосомы

Такое представление решения дает возможность использовать оператор рекомбинации наборов хромосом. Этот оператор осуществляет перенос отдельно взятой хромосомы из одного решения в другое. Также мультихромосомное представление решения дает возможность поиска, при котором некоторые отдельные хромосомы остаются неизменными.

*Оператор отбора.*

Для реализации оператора отбора был выбран метод рулетки. При этом каждая особь, представляющая определенное решение, имеет поле пропорциональное ее значению функции приспособленности. Для отбора одной родительской мультихромосомы колесо рулетки вращается, и после останова ее указатель определяет искомую особь. Очевидно, чем больше приспособленность особи, представляемой мультихромосомой, тем больше вероятность выбора ее в качестве родительской. На рис. 3 изображен пример колеса рулетки для популяции из пяти особей.

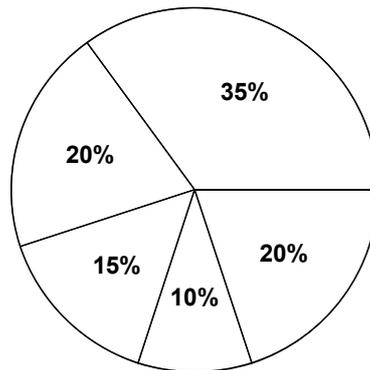


Рисунок 3 — Колесо рулетки оператора отбора

В популяции  $\Psi$  для каждой особи  $\psi_i$  вероятность того, что она будет отобрана при одном запуске оператора, определяется выражением:

$$P(\psi_i) = \frac{T(\psi_i)}{\sum_{\forall \psi_i \in \Psi} T(\psi_i)} \tag{2}$$

*Оператор рекомбинации.*

Для реализации оператора рекомбинации отбираются две особи. Затем каждая отдельная хромосома в мультихромосоме одной особи с вероятностью  $P_{рек}$  меняется местами с соответствующей хромосомой другой особи (рис. 4).

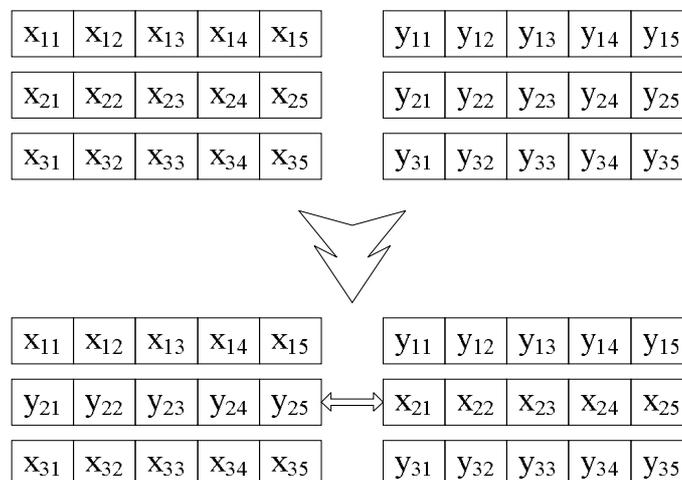


Рисунок 4 — Оператор рекомбинации

Алгоритм реализации оператора рекомбинации:

1. Выбрать с помощью оператора отбора две особи:  $\psi_1$  и  $\psi_2$  с матрицами распределения данных  $A_1$  и  $A_2$  соответственно.
2. Для каждой из пар строк имеющих одинаковые порядковые номера матриц  $A_1$  и  $A_2$  с вероятностью  $P_{рек}$  выполнять их обмен.
3. В результате получаем две особи представленные матрицами  $A'_1$  и  $A'_2$ .

Рассмотрим принцип работы оператора рекомбинации на примере.

Для особей, представленных матрицами  $A_1$  и  $A_2$  которые имеют вид:

$$A_1 = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix} \text{ и } A_2 = \begin{bmatrix} 1 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 1 \end{bmatrix},$$

рекомбинация хромосом представленных вторыми строками приведет к созданию двух новых особей с матрицами распределения данных  $A'_1$  и  $A'_2$ :

$$A'_1 = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix} \text{ и } A'_2 = \begin{bmatrix} 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 \end{bmatrix}.$$

*Оператор скрещивания.*

После оператора рекомбинации применяется оператор скрещивания. Для этого с помощью оператора отбора выбираются две родительские особи. Затем каждая пара хромосом из мультихромосом родительских особей с вероятностью  $P_{скреци}$  обмениваются своими частями.

При этом хромосомы представляются в виде циклических последовательностей, когда последний элемент соединяется с первым. Случайным образом выбираются две точки разрыва хромосом. По этим точкам хромосомы делятся на две части. Затем происходит обмен соответствующими частями между хромосомами (рис. 5)

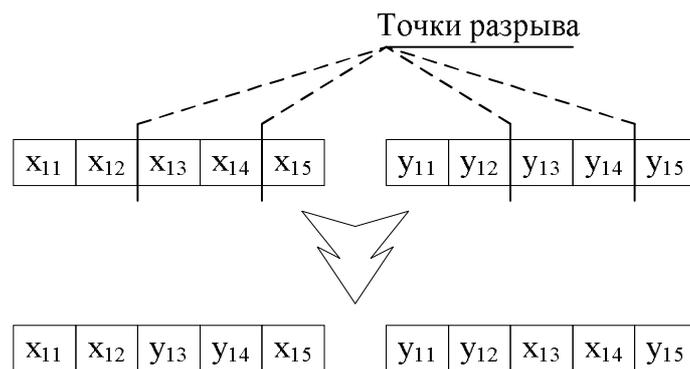


Рисунок 5 — Оператор скрещивания

Алгоритм выполнения оператора скрещивания:

1. Выбрать с помощью оператора отбора две особи:  $\psi_1$  и  $\psi_2$  с матрицами распределения данных  $A_1$  и  $A_2$ .
2. Для каждой из пар строк имеющих одинаковые порядковые номера матриц  $A_1$  и  $A_2$  с вероятностью  $P_{скреци}$  выполнять 2.1 – 2.2.

- 2.1. Случайным образом определяем две точки разрыва хромосом:  $b_1 \in [1, n]$  и  $b_2 \in [1, n]$ ,  $(b_1 \neq b_2)$ .
- 2.2. Выполнить обмен участками хромосом, получаемыми по точкам разрыва.
3. В результате получаем две особи представленные матрицами  $A'_1$  и  $A'_2$ .

Проиллюстрируем принцип работы оператора скрещивания на примере двух особей, представленных матрицами  $A_1$  и  $A_2$  которые имеют вид

$$A_1 = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix} \text{ и } A_2 = \begin{bmatrix} 1 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 1 \end{bmatrix}.$$

В результате скрещивания хромосом, представленных вторыми строками с точками разрыва  $b_1 = 3$ ,  $b_2 = 5$  будут получены две новые особи представленные матрицами  $A'_1$  и  $A'_2$ :

$$A'_1 = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix} \text{ и } A'_2 = \begin{bmatrix} 1 & 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 \end{bmatrix}.$$

*Оператор мутации.*

После того как, с помощью операторов рекомбинации и скрещивания, сформирована новая популяция, к ее членам применяется оператор мутации. Оператор мутации заключается в инвертировании элементов хромосом с определенной, обычно очень маленькой, вероятностью.

При этом с вероятностью  $P_{мут}$  каждое значение матрицы распределения данных  $A$  меняется на противоположное.

При применении операторов ГА необходимо каждое новое решение проверять на соответствие ограничениям.

**Алгоритм оптимизации распределения данных**

Исходной информацией для алгоритма оптимизации является: количество поколений ГА; размер популяции (количество особей в популяции); количество хромосом в одной мультихромосоме; размер хромосомы (количество ген в хромосоме); Вероятности применения операторов ГА (рекомбинации, скрещивания, мутации).

После того как исходная информация для алгоритма оптимизации введена, начинается формирование начальной популяции ГА. Для этого случайным образом, но с проверкой на соответствие ограничениям, формируются мультихромосомы  $N_p$  особей.

Когда начальная популяция сформирована, начинается основной цикл алгоритма, в котором к особям популяции применяются операторы ГА.

Первым из этих операторов является оператор рекомбинации. Для рекомбинации, с помощью оператора отбора выбираются две родительские особи  $\psi_1$  и  $\psi_2$ . Затем в цикле перебираются пары хромосом этих особей и с вероятностью  $P_{рек}$  эти хромосомы меняются местами. В результате получаются две особи  $\psi'_1$  и  $\psi'_2$ . Эти особи проверяются на принадлежность области допустимых решений. Если полученные решения не удовлетворяют ограничениям, то происходит их корректировка. Затем вычисляется значение функции приспособленности полученных особей. После этого две лучшие особи из пар родителей и потомков добавляются в новую популяцию.

С помощью оператора рекомбинации происходит формирование промежуточной популяции. Новая популяция формируется на основе этой промежуточной с помощью

оператора скрещивания аналогичным образом. Отличие оператора скрещивания от рекомбинации заключается в том, что происходит обмен не целыми хромосомами, а их частями. Для этого случайным образом выбираются две точки разрыва хромосом. По этим точкам хромосомы делятся на две части, затем они обмениваются этими частями. Оператор скрещивания к каждой паре хромосом применяется с вероятностью  $P_{скрещ}$ .

Когда все особи новой популяции сформированы, к ним применяется оператор мутации. Он заключается в инвертировании отдельных генов особи.

Основной цикл алгоритма повторяется определенное количество раз. При этом каждое последующее поколение в целом получается лучше предыдущего (по крайней мере не хуже, так как дочерние особи попадают в новую популяцию, только если они лучше родительских). После того как отработано последнее поколение  $G$  процесс оптимизации заканчивается. Лучшее из полученных решений является субоптимальным.

### **Результаты экспериментальной проверки метода**

Для проверки эффективности разработанного алгоритма оптимизации РБД, с использованием модифицированного генетического алгоритма и объектной модели, проведен ряд вычислительных экспериментов и выполнена статистическая обработка полученных результатов. В качестве объекта экспериментальных исследований выбрана компьютерная информационная система ЗАО ПО “Киев-Конті”. Для данного объекта с помощью разработанного метода оптимизации было получено субоптимальное значение критерия эффективности, абсолютное отклонение которого от глобального минимума составляет 3,09 с, относительное отклонение — 3,76%. При этом, уменьшение суммарного среднего времени выполнения запросов и распространения обновлений по сравнению со значением, полученном при исходном распределении данных 26,47%.

### **Заключение**

Разработанный метод оптимизации РБД позволяет находить варианты распределения данных, при которых суммарное среднее время выполнения запросов и распространения обновлений, меньше чем при исходном распределении данных, за счет приближения данных к местам их наиболее интенсивного использования и определения оптимального числа копий данных на различных узлах. Таким образом, совместное использование объектной модели распределенных баз данных и модифицированного генетического алгоритма позволяет находить варианты субоптимального распределения данных, при которых достигается повышение производительности РБД без дополнительных капитальных затрат на модернизацию оборудования.

### **Литература**

1. Цегелик Г.Г. Системы распределенных баз данных. — Львов.: Свит, 1990. — 168 с.
2. Мамиконов А.Г., Кульба В.В., Косяченко С.А., Ужастов И.А. Оптимизация структур распределенных баз данных в АСУ. — М.: Наука 1990.— 240 с.
3. Лаздынь С.В., Телятников А.О. Новый подход к оптимизации распределения данных в компьютерных информационных системах // Тези доповідей учасників VI Міжнародної науково-практичної конференції “Системний аналіз та інформаційні технології”. — Київ: НТУУ “КПІ”, 2004. — С. 222–224.
4. Лаздынь С.В., Телятников А.О. Повышение эффективности распределенных баз данных с использованием объектно-ориентированного моделирования и генетических алгоритмов // Единое информационное пространство: Сб. докл. Междунар. научно-практич. конф. — Днепропетровск: ИПК ИнКомЦентра УГХТУ, 2003. — С. 23–26.
5. Божич В.И., Лебедев В.Б. Методы генетического поиска для решений представимых мультихромосомами // Перспективные информационные технологии и интеллектуальные системы. — 2002. — № 3. — С. 38–44.