

*Грицук А.С. — рецензент Воевода А.А.
УНК “ИПСА” НТУУ “КПИ”, Киев, Украина*

Селекция методом рулетки в генетических алгоритмах при оптимизации запросов к базам данных

На сегодняшний день базы данных являются основным средством для хранения больших объемов структурированных данных. Количество генерируемой человечеством информации растет ежегодно, и, в связи с этим, растут объемы баз данных. Также базы данных используются для хранения информации, доступ к которой организуется через Интернет.

Одной из главных задач в проектировании баз данных на сегодняшний день является их оптимизация. Одним из направлений является оптимизация запросов. Она состоит в выборе оптимального порядка перебора таблиц. Существует несколько алгоритмов решения этой задачи, в том числе – генетические алгоритмы.

Ключевым этапом генетического алгоритма является селекция. Существует несколько подходов к организации селекции, но базовым и наиболее популярным считается так называемый метод рулетки (roulette-wheel selection).

Функционирует данный алгоритм следующим образом. Каждой хромосоме может быть сопоставлен сектор колеса рулетки, величина которого устанавливается пропорциональной значению функции приспособленности данной хромосомы по формуле

$$P_{sel}(i) = \frac{f(i)}{\sum_{i=1}^n f(i)}$$

Далее генерируется соответствующее количество случайных чисел. Хромосомы, в сектора которых входят эти числа, считаются отобранными.

При таком отборе члены популяции с высшей приспособленностью будут чаще выбираться, чем особи с низкой приспособленностью. Такой подход, с одной стороны, обеспечивает отбор наиболее подходящих хромосом, а с другой – оставляет возможность выбора менее приспособленных, которые все же могут быть полезны.

Данный алгоритм является самым распространенным среди методов селекции, поскольку предлагает естественный и достаточно эффективный метод отбора, при этом являясь наиболее простым в программной реализации, и не требует больших вычислений в отличие от главной альтернативы – метода турнирного отбора.

Данный алгоритм обеспечивает оптимальное соотношение количества перестановок таблиц в базе данных к объему вычислений самого алгоритма, обеспечивает возможность рассмотрения неоптимальных путей. Таким образом, мы можем найти лучшие локальные минимумы функции стоимости, по сравнению с которыми глобальный минимум не дает существенного преимущества.

Литература

1. Гарсиа-Молина Г., Ульман Д., Уидом Д. “Системы баз данных. Полный курс” – М.: Изд Вильямс, 2003. – 1088 с.
2. D. Knuth. “The Art of Computer Programming, Volume 4, Fascicle 0: Introduction to Combinatorial Algorithms and Boolean Functions” – Addison-Wesley, 2008. – 240 с.
3. Дюк В., Самойленко А. Data Mining: учебный курс (+CD). – СПб: Изд. Питер, 2001. – 368 с.
4. Генетические алгоритмы[Электронный ресурс]/. – Режим доступа – www. URL: <http://www.itfru.ru/index.php/genetic-algorithms> – 20.02.2011.