

Раздел III. Алгоритмическое и программное обеспечение

УДК 681.142

А.М. Белевцев, М.А. Дружинин

ОБЩАЯ СТРУКТУРА КОМБИНИРОВАННОГО ЭВОЛЮЦИОННО-ГЕНЕТИЧЕСКОГО АЛГОРИТМА ПОИСКА РЕШЕНИЙ В ЗАДАЧАХ ОПТИМИЗАЦИИ ИНФОРМАЦИОННО-ВЫЧИСЛИТЕЛЬНЫХ ПРОЦЕССОВ

Генетический алгоритм – это специальный алгоритм поиска, используемый для решения задач оптимизации и моделирования путем последовательного подбора, комбинирования и вариации искомых параметров с использованием механизмов, напоминающих биологическую эволюцию. Предложена следующая структура генетического алгоритма: создание начальной популяции; определение (задание) функций приспособленности для особей популяции (оценивание); выбор индивидов из текущей популяции (селекция); скрещивание и/или мутация; вычисление функций приспособленности для всех особей; формирование нового поколения.

Количество итераций генетического алгоритма находится в большой зависимости от «качества» начальной популяции. При создании начальной популяции большую роль играют подходы и методы эвристического плана. Построение эвристического алгоритма формирования начальной популяции для типовых задач целераспределения основывается на использовании матрицы возможных назначений.

Генетические алгоритмы оптимизации; начальная популяция; размножение; скрещивание; компактные строительные блоки.

А.М. Belevtsev, M.A. Druzhinin

THE GENERAL STRUCTURE OF THE COMBINED EVOLUTIONARY AND GENETIC ALGORITHM OF SEARCHING THE DECISIONS IN THE OPTIMIZATION INFORMATION PROCESSES

The genetic algorithm – is the special algorithm of search used for the solution of problems of optimization and modeling by consecutive selection, a combination and a variation of required parameters with the use of the mechanisms reminding biological evolution. The following structure of genetic algorithm is offered: creation of initial population; definition (task) of functions of fitness for individuals of population (estimation); choice of individuals from the current population (selection); crossing and/or mutation; fitness functions evaluation for all individuals; formation of new generation.

The number of iterations of genetic algorithm depends on the "quality" of initial population. In the creation of initial population there is a big role of the methods of the heuristic plan. Creation of heuristic algorithm of the formation of initial population for standard problems of target distribution is based on the use of matrix of possible appointments.

Genetic algorithms of optimization; initial population; reproduction; crossing; compact construction blocks.

Генетический алгоритм – это специальный алгоритм поиска, используемый для решения задач оптимизации и моделирования путем последовательного подбора, комбинирования и вариации искоемых параметров с использованием механизмов, напоминающих биологическую эволюцию. Отличительной особенностью генетического алгоритма является акцент на использование оператора «скрещивания», который производит операцию рекомбинации решений-кандидатов, роль которой аналогична роли скрещивания в живой природе.

Общим представлением задач переборного типа, к которым относятся и генетические алгоритмы оптимизации (ГАО), являются модели математического целочисленного программирования. Чаще всего в качестве независимых переменных в таких моделях фигурируют булевы переменные 0 и 1. Такие модели часто называют моделями псевдобулевого математического программирования. Такое название связано с тем, что в указанных моделях булевыми являются лишь переменные, в то время как целевая функция и ограничения выражаются рациональными значениями [1].

Генетические алгоритмы не критичны к размерности задачи и виду целевой функции, которая может быть нелинейной, разрывной, недифференцируемой и многоэкстремальной. Неявный параллелизм ГАО позволяет ему тестировать и использовать большое количество областей в пространстве решений, работая с относительно небольшим числом цепочек.

Оптимизационная задача ГАО кодируется таким образом, чтобы её решение могло быть представлено в виде вектора («хромосомы»). Некоторым образом создаётся определённое количество начальных векторов («начальная популяция»). Они оцениваются с использованием «функции приспособленности», в результате чего каждому вектору присваивается определённое значение («приспособленность»), которое определяет вероятность выживания организма (участия полученного решения в новых итерациях), представленного данным вектором. После этого с использованием полученных значений приспособленности выбираются вектора (селекция), допущенные к «скрещиванию». К этим векторам применяются «генетические операторы» (в большинстве случаев «скрещивание» и «мутация»), создавая таким образом следующее «поколение». Особи следующего поколения также оцениваются, затем производится селекция, применяются генетические операторы и т.д. Так моделируется «эволюционный процесс», продолжающийся несколько жизненных циклов (поколений), пока не будет выполнен критерий остановки алгоритма. Таким критерием может быть:

- ◆ нахождение глобального либо субоптимального решения;
- ◆ исчерпание числа поколений, отпущенных на эволюцию;
- ◆ исчерпание времени, отпущенного на эволюцию.

Как показывают имеющиеся исследования, эффективность генетических алгоритмов во многом определяется структурой и составом начальной популяции. Поэтому при создании начальной популяции следует использовать знания о физической сущности и специфике решаемой задачи. Эти знания могут базироваться на опыте разработчика, сведениях об аналогичных разработках, литературных источниках и т.д. Иными словами, при создании начальной популяции большую роль играют подходы и методы эвристического плана [2].

Построение эвристического алгоритма формирования начальной популяции для типовых задач оптимизации систем с распределённой обработкой данных (СРОД) существенным образом ориентировано на учёт специфики информационных взаимодействий в СРОД в плане сокращения общего объёма обрабатываемых узлов и динамической информации. Использование матрицы возможных назначений (МВН) позволяет наглядно отобразить физическую сущность процедуры поиска решения в модели псевдобулевого математического программирования.

Если задача малой размерности или МВН слабо заполненная (относительно «1»), то существует вероятность получения оптимального решения уже в рамках построения начальной популяции. В противном случае начальная популяция используется как исходная информация для последующих процедур генетического алгоритма. При этом ожидается значительное сокращение времени, затраченного ГАО на решение данной задачи.

С учётом изложенного существование любого генетического алгоритма может быть отображено путём моделирования основных факторов эволюционного развития популяции по следующей обобщённой схеме [3, 4, 5]:

1. Формирование начальной популяции из n особей.

Заключается в генерации хромосомного набора из K бинарных строк, удовлетворяющих требованиям, предъявляемым к математической модели оптимизационной задачи, и вычислении степени приспособленности особей (определяемых бинарными строками) к внешней среде (целевой установке задачи).

2. Воспроизводство "потомков" с наследными признаками "родителей".

Осуществляется реализация следующих основных процедур:

- ◆ выбор конкретной "родительской" пары для участия в процессе размножения (процедура скрещивания); наиболее известны процедуры скрещивания по схеме панмиксии (случайное скрещивание), инбридинга и аутбридинга, ассортативного скрещивания;
- ◆ выбор схемы размножения (размножение по схеме рекомбинации генов, простого или сложного кроссинговера);
- ◆ построение по выбранной схеме из генотипов "родителей" генотипов "потомков", сохраняющих наследственные признаки "родителей" с последующим вычислением степени приспособленности "потомков" к внешней среде.

3. Проведение операции мутагенеза, приводящей к генетическим изменениям родительских генотипов, заключается в осуществлении следующих действий:

- ◆ выбор типа мутагенеза (точечная мутация, генная мутация, макромутация, хромосомная мутация);
- ◆ построение по генотипу одной из особей популяции генотипа особи "мутанта" с помощью выбранного типа мутации.

4. Осуществление естественного отбора.

Выполняется по следующей схеме:

- ◆ определение среди "родителей", "потомков" и "мутантов" особей, образующих репродуктивную группу, которая принимает участие в естественном отборе (известен ряд схем формирования репродуктивных групп: общая, элитарная, селекционная и др. схемы);
- ◆ выбор схемы естественного отбора (возможны "жёсткая" либо "мягкая" схемы отбора);
- ◆ формирование по выбранной схеме хромосомного отбора популяции следующего поколения из особей репродуктивной группы.

Таким образом, исходной информацией для ГАО является генетический код, представляющий структуру исследуемых процессов. Поиск оптимальных решений заключается в поиске двоичных цепочек из всего их многообразия, в наибольшей степени удовлетворяющих заданному критерию поиска. Отдельные области в пространстве решений определяются по наличию нулей и единиц в определённых позициях цепочек. Примером такой области является множество цепочек, начинающихся с комбинации 1011...

Генетические алгоритмы покрывают функциональный "ландшафт" задачи сетью. Множество цепочек в развивающейся популяции одновременно тестирует множество областей. При этом частота, с которой генетический алгоритм тестирует

ет различные области, пропорциональна близости к оптимальному решению, что повышает вероятность отыскания такого решения. Эта способность ГАО концентрировать внимание на наиболее перспективных областях пространства решений является прямым следствием их способности комбинировать цепочки, содержащие частные решения. Сначала оценивается каждая цепочка популяции, чтобы определить "степень" оптимальности кодируемого ею решения. Далее получившие наилучшие оценки цепочки спариваются. При этом две цепочки располагаются друг перед другом, случайным образом выбирается точка, и части, расположенные слева от этой точки, меняются местами, образуя двух потомков: один из них содержит символы первой цепочки вплоть до точки перекрещивания, а далее – символы другой. Другой потомок содержит дополнительный набор символов. Такой процесс (оператор) называется кроссинговером. После этого потомки возвращаются в популяцию, но не замещают родителей, а замещают цепочки с наихудшей оценкой кодируемого ими решения. Последние отбрасываются в каждом поколении, так что численность популяции со временем остаётся постоянной.

Наличие механизма скрещивания цепочек усложняет эффект неявного параллелизма. Цель скрещивания в ГАО заключается в том, чтобы протестировать новые области, вместо того, чтобы тестировать одну и ту же цепочку в новых поколениях. Этот процесс может переместить потомство из одной области в другую, в результате чего частота тестирования различных областей отклоняется от строго пропорциональной зависимости от среднего показателя. Это отклонение замедляет процесс эволюции.

Близко соседствующие единицы и нули, определяющие пределы области, называются компактными строительными блоками (КСБ). У этих блоков наибольшая вероятность сохраниться после кроссинговера в неизменном виде и, таким образом, распространиться в будущих поколениях. Хотя механизм размножения, включающий скрещивание, не даёт возможности тестировать все области с частотой, пропорциональной их качеству, он обеспечивает её для областей, которые определяются КСБ. Количество же КСБ в популяции цепочек намного превосходит число самих цепочек, благодаря чему ГАО проявляет свойства неявного параллелизма.

Другим важным оператором ГАО является оператор мутации. Этот оператор изменяет небольшой участок сети: примерно один на каждые 10 000 символов переключается с 0 на 1 и наоборот. Сами по себе мутации обычно не дают прогресса в поиске решения, но они страхуют от возникновения однородной популяции, не способной к дальнейшей эволюции.

Ещё один важный оператор ГАО – оператор инверсии. Операция, называемая в генетике живых организмов инверсией, время от времени меняет расположение генов таким образом, что гены, расположенные у родительских цепочек далеко друг от друга, могут стать соседями в потомстве. В определённом смысле инверсия является разновидностью мутации.

Инверсия равносильна переопределению КСБ, в результате которого он становится более компактным и менее уязвимым для кроссинговера. При этом более компактная версия блока замещает менее компактную, поскольку теряет меньше потомства в результате ошибок, возникающих при воспроизведении. В результате система адаптации, использующая инверсию, отдаёт предпочтение компактным версиям полезных блоков.

Таким образом, предложена общая структура комбинированного эволюционно-генетического алгоритма поиска решений в задачах оптимизации информационно-вычислительных процессов. В соответствии с выбранной структурой проводятся исследования, которые позволят определить рациональные варианты построения основных этапов ГАО и их эффективность в зависимости от «качества» начальной популяции.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Балыбердин В.А., Белевцев А.М., Степанов О.А.* Оптимизация информационных процессов в автоматизированных системах с распределённой обработкой данных. – М.: Технология, 2002.
2. *Домбровский Я.Е., Дружинин М.А.* О построении начальной популяции решений в задачах оптимизации технологических процессов в распределённых системах запросного типа: Сборник статей «Гагаринские чтения». – М., 2009. – 2 с.
3. *Davis L.* Handbook of Genetic Algorithms. – N.J., van Nostrand, 1991. – 408 p.
4. *Батищев Д.И.* Генетические алгоритмы решения экстремальных задач. – Воронеж: Изд-во ВГТУ, 1995. – 132 с.
5. *Гладков Л.А., Курейчик В.В., Курейчик В.М.* Генетические алгоритмы. – М.: Физматлит, 2006. – 318 с.

Статью рекомендовал к опубликованию д.т.н., профессор В.А. Петраков.

Белевцев Андрей Михайлович – Информационно-аналитический инновационный центр МАТИ; e-mail: ambelevtsev@yandex.ru; 121522, г. Москва, ул. Оршавская, 3; тел.: 89037691788; д.т.н.; профессор; научный руководитель.

Дружинин Михаил Александрович – 3 Центральный научно-исследовательский институт Министерства обороны Российской Федерации; e-mail: ambelevtsev@yandex.ru; 107564, Москва, Погонный пр-д, 10; тел.89261334779; начальник лаборатории.

Belevtsev Andrey Michailovich – Information-Analytical Innovation Center MATI; e-mail: ambelevtsev@yandex.ru; 3, Orshavckay street, Moscow, 121522, Russia; phone: +79037691788; dr. of eng. sc.; professor; science supervisor.

Druzhinin Mihail Aleksandrovich – 3 The central scientific research institute of the Ministry of Defence of the Russian federation; e-mail: ambelevtsev@yandex.ru; 10, Pogonny'j, Moscow, 107564, Russia; phone: +79261334779; a laboratory head.

УДК 519.688

С.Л. Беляков, М.Л. Белякова, И.Н. Розенберг

**ОГРАНИЧЕНИЯ ЦЕЛОСТНОСТИ ПРИ ВИЗУАЛИЗАЦИИ
ПРОСТРАНСТВЕННОЙ БАЗЫ ДАННЫХ***

Работа посвящена проблеме визуализации информации, содержащейся в базах пространственных данных информационных систем. Необходимость автоматической визуализации ставит задачу построения формальных соотношений для оценки целостности базы данных на уровне представления в виде карты или схемы. В статье проанализированы особенности концептуальной модели визуализации. Предложены оценочные булевы функции, значения которых используются для выявления аномалий добавления и удаления элементов данных в карту или схему. Описаны условия возникновения дефектов отображения поисковых схем. Сформулированы принципы учета топологических отношений между элементами визуального представления базы данных. Приведены соотношения для оценки целостности с учетом топологии пространственных объектов. Предложена нечеткая модель классификации ограничений целостности.

Визуализация баз данных; геоинформационные системы; целостность.

* Работа выполнена при финансовой поддержке РФФИ, проекты №11-01-00011 и №12-01-00032.