Н.Г. Графеева,
Л.И. Григорьева,
Н.Л. Калинина-Шувалова

Генетические алгоритмы и теория Дарвина: к проблеме применения

Genetic algorithms and Darwin's theory: the problems of the application

Графеева Наталья Генриховна
Санкт-Петербург, Россия
Доцент Санкт-Петербургского государственного университета, кандидат физико-математических наук,
N.Grafeeva@spbu.ru

Grafeeva Natalia H.
Saint-Petersburg, Russia
Professor of Saint-Petersburg State University
Ph.D., Candidate of Physical and Mathematical Sciences
N.Grafeeva@spbu.ru

Григорьева Людмила Ивановна
Санкт-Петербург, Россия
Старший преподаватель Санкт-Петербургского государственного университета,
l.grigorieva@spbu.ru

Grigorieva Lyudmila Ivanovna
Saint-Petersburg, Russia
Senior Teacher of Saint-Petersburg State University
l.grigorieva@spbu.ru

Калинина-Шувалова Наталия Леонидовна
Санкт-Петербург, Россия
Старший преподаватель Санкт-Петербургского государственного университета,
n.kalinina-shuvalova@spbu.ru

Kalinina-Shuvalova Natalie Leonidovna
Saint-Petersburg, Russia
Senior Teacher of Saint-Petersburg State University
n.kalinina-shuvalova@spbu.ru

Аннотация: Конкуренция – один из движущих механизмов развития. Этот механизм поистине всеохватен. Конкуренция работает в экономике и позволяет создавать все более совершенные технологии и новые товары. Но экономика – только частный случай того, как в результате отбора получается наилучший результат. Генетические алгоритмы, в основе которых лежит конкурентная борьба особей, являются одними из интереснейших подходов к поиску решений для самых разнообразных задач. Статья посвящена соотнесению особенностей генетических алгоритмов и основных положений теории естественного отбора. В качестве примеров, которые решаются с помощью генетических алгоритмов, выбраны известные алгебраические задачи: нахождение положительных корней Диофантовых уравнений, представление Лагранжа для натуральных чисел, поиск чисел Армстронга.

Summary: Competition is one of the drivers of development. This mechanism is truly inclusive. Competition works in the economy and allows you to create more sophisticated technologies and new products. But the economy presents only a special case of best results in selection process. Genetic algorithms which are based on competition of individuals, are one of the most interesting methods to look for solutions for different tasks. The article is devoted to the correlation of the features of genetic algorithms and the basic provisions of natural selection theory. As examples for the use of genetic algorithms some problems of number theory were selected: looking for Diophantine equations roots, Lagrange's four square theorem, Armstrong's numbers.

Ключевые слова: экономика, конкуренция, генетический алгоритм; теория естественного отбора; фитнес-функция; разложение Лагранжа; числа Армстронга; Диофантовы уравнения.

Keywords: economy, competition, enetic algorithm; the theory of natural selection; fitness function; Lagrange's four square theorem; Armstrong's numbers; Diophantine equations.

Генетические алгоритмы привлекают к себе внимание уже не первое десятилетие. К настоящему времени существует множество литературы, в которой описываются как обобщенные принципы построения таких алгоритмов, так и конкретные примеры, призванные решать самые разнообразные задачи. Одними из первых публикаций считаются работы

Голланда (Holland). Классикой в этой области стала книга «Адаптация в естественных и искусственных системах» («Adaptationin Natural and Artifical Systems», 1975) [8].

Генетические алгоритмы взяли за свою основу идеи Чарльза Дарвина, которые он опубликовал в 1859 году в своей работе «О происхождении видов путем естественного отбора или сохранении благоприятствуемых пород в борьбе за жизнь» [4].

Существует большое количество задач, при решении которых могут использоваться генетические алгоритмы. Рассмотрим несколько задач, чтобы на примере показать, как именно понятия теории естественного отбора находят свое отражение в построении генетических алгоритмов.

В качестве первого примера выбрано представление Лагранжа для натурального числа. Оно дает разложение числа на сумму квадратов четырех неотрицательных целых чисел, т.е:

$$a_1^2 + a_2^2 + a_3^2 + a_4^2 = A$$

Например:

$$1^2 + 1^2 + 1^2 + 3^2 = 12$$

$$0^2 + 0^2 + 3^2 + 3^2 = 18$$

Доказано, что такое представление существует для любого натурального числа.

Второй пример – числа Армстронга. Натуральное число считается числом Армстронга, если выполняется следующее равенство:

$$a_1^n + a_2^n + \cdots + a_n^n = A,$$

где A – исходное число, n – количество цифр в A, a_i – i-ая цифра числа A. Примером такого числа является 153. В этом случае n=3.

$$1_1^n + 5_2^n + 3_n^n = 1 + 125 + 27 = 153$$

Третий пример – линейное Диофантово уравнение с четырьмя неизвестными:

$$a_1x + a_2y + a_3z + a_4u = A$$

Вот один из примеров Диофантова уравнения:

$$2x + y + 3z + 7u = 54$$

Далее будем рассматривать только целые положительные корни Диофантова уравнения и положительные коэффициенты.

Посмотрим на каждое из положений теории генетического отбора и попытаемся трактовать его в терминах генетических алгоритмов вообще и для вышеперечисленных примеров в частности.

Популяция любого вида состоит из различающихся особей.

В природе особь любого вида уникальна, будь то кошка или человек. Вы не найдете двух абсолютно одинаковых людей или бабочек.

С точки зрения генетического алгоритма популяция представляет собой множество элементов, среди которых ищется решение задачи. Элемент может иметь простой тип (например, целое число, дата, строковое значение), либо представлять собой структуру, состоящую из многих компонентов (например, вектор из натуральных чисел). Каждый элемент множества – особь популяции. Решение задачи – элемент множества и также одна из особей популяции.

В случае разложения Лагранжа особь представляет собой 4-х-мерный вектор с неотрицательными целыми координатами. Множество особей, среди которых ищется решение, не зависит от числа, для коего происходит поиск представления. Более того, с точки зрения генетического алгоритма, особи (1, 1, 1, 9) и (1, 1, 9, 1) можно считать не различающимися. Для этого будем рассматривать только вектора, у которых координаты упорядочены по возрастанию. Тем самым количество рассматриваемых особей можно значительно сократить.

Для чисел Армстронга множество особей ограничено размерностью n, которая является параметром алгоритма. Для n=3 поиск будет происходить среди 3-разрядных натуральных чисел, для n=4 – среди 4-разрядных и т.д. В каждом случае искомое число соответствующей размерности можно, как и в случае представления Лагранжа, считать 3-хмерным вектором, 4-хмерным вектором

или вектором размерности n.Здесь порядок координат является существенным. Хотя с математической точки зрения 153 и 135 дадут арифметические выражения, различающиеся только порядком слагаемых, но в первом случае сумма будет совпадать с исходным числом (153), а во втором – нет (сумма по-прежнему будет 153, а не 135).

При решении Диофантовых уравнений ответ также ищется среди векторов с натуральными координатами. В данном случае размерность вектора совпадает с количеством переменных в уравнении. Порядок координат также важен, т.к. каждая из координат должна умножаться на соответствующий коэффициент уравнения.

Особи вида конкурируют между собой за ресурсы.

В природе особи конкурируют между собой за пищу, за территорию, на которой пищу можно добывать, за лучшую самку или самца, чтобы иметь наиболее жизнеспособное потомство.

Данный принцип можно рассматривать с точки зрения генетических алгоритмов как стремление особи быть решением задачи.

Новых особей рождается больше, чем остается в результате борьбы за ресурсы

Природа позволяет появиться большему количеству детенышей, чем может существовать и / или прокормиться в той или иной местности. В дальнейшем работает естественный отбор, чтобы количество особей свелось к оптимальному. Особи, наименее приспособленные к окружающему миру, погибают.

В генетическом алгоритме требуется замена естественному отбору. Нужен процесс, который бы имитировал борьбу особей за ресурсы и выбраковывал наименее приспособленные особей, т.е. особей, которые в меньшей степени соответствуют тому, чтобы быть решением исходной задачи. Ведь именно способность быть решением мы и рассматриваем в качестве основной цели в борьбе за выживание особей при работе генетического алгоритма.

В качестве аналога процесса естественного отбора в генетических алгоритмах выступает фитнес-функция. Фитнес-функция вычисляет «близость» особи к тому, чтобы быть решением рассматриваемой задачи. Для всех вышеупомянутых примеров фитнес-функция вычисляется похожим способом, а именно: значения координат вектора особи подставляется в Диофантово уравнение, разложение Лагранжа, формулу для проверки чисел Армстронга. Далее определяется разность между полученным значением и требуемым. Чем модуль разности меньше, тем более хорошей (более приспособленной для выживания) считается особь.

$f = A - (a_1^2 + a_2^2 + a_3^2 + a_4^2)$ _ фитнес – функция для разложения Лагранжа

$f = A - (a_1^n + a_2^n + \ldots + a_n^n)$ _ фитнес – функция для чисел Армстронга

$f = A - (a_1 x + a_2 y + a_3 z + a_4 u)$ _ фитнес – функция для Диофантова уравнения.

В естественной среде нет понятия начального, самого первого поколения. Но для генетического алгоритма приходится искусственно создавать такое поколение. Дальнейшие популяции строятся на основе начального поколения.

При построении начального поколения для генетического алгоритма можно породить особи случайным образом и положиться на то, что далее генетический алгоритм обеспечит построение решения.

Особи, имеющие качество(а), помогающие в борьбе за ресурсы, выживают успешнее и имеют потомство, которому могут передать это качество(а).

В естественной среде особи, имеющие лучшие показатели в определенных сферах, выживают успешнее и чаще имеют потомство. За те или иные характеристики особи несут ответственность ее гены, которые могут передаваться по наследству. Поскольку слабые особи реже имеют потомство, то их генетический набор реже может быть унаследован последующими поколениями.



В процессе работы генетического алгоритма также происходит воспроизводство особей для очередного поколения. При этом каждая из особей, являющихся родителем, предоставляет свой генетически набор. Из двух наборов формируется комплект особи нового поколения. В качестве генетических наборов для рассматриваемых нами примеров выступают координаты векторов. Для особи-потомка выбирается одно из двух значений координат, предоставляемых родителями. Так происходит для каждой координаты.

Рассмотрим процесс немного подробнее. На входе каждого витка имеется *начальное* количество особей поколения. Далее особи начинают бороться за ресурсы, каждая из особей рассматривается с точки зрения пригодности быть решением задачи, т.е. для каждой особи вычисляется фитнес-функция. Особи, наименее подходящие для того, чтобы быть решением, погибают, генетический алгоритм признает нецелесообразным их использование в следующем поколении. Особи, более близкие к тому, чтобы быть решением задачи, переходят в следующее поколение или используются для генерации особей следующего поколения.

Качество, характеризующее особь, может либо передаться по наследству от одного из родителей, либо возникнуть в результате мутации, произошедшей по какой-либо внешней причине.

Кроме обмена генами при скрещивании особей, в естественной среде случаются мутации. При мутациях один или несколько генов могут получить повреждения. Фактически, при этом получается расширение генетического набора для поколения. Если мутировавшая особь более приспособлена для выживания, то ее комплект будет передаваться по наследству следующим поколениям.

При работе генетических алгоритмов также применяется понятие мутации. В тех примерах, которые рассматриваются в статье, при мутации особи одна или несколько координат вектора могут быть заменены на случайные значения. Таким образом, если свершившаяся мутация ведет к появлению более приспособленной для жизни особи (более близкой к тому, чтобы быть решением), то особь с большой вероятностью войдет в число родителей для следующего поколения или вообще сама перейдет в следующее поколение. Если же мутация неудачна, то особь останется за бортом процесса воспроизводства.

Таким образом, мы рассмотрели все положения теории естественного отбора. Далее остановимся еще на некоторых аспектах генетических алгоритмов применительно к выбранным примерам.

Некоторые особенности выбора особей для следующего поколения.

Процесс работы генетического алгоритма представляет собой цикл. На каждом шаге цикла просматриваются особи текущего поколения. Часть из них выбирается в качестве родителей для следующего поколения, часть переходит в следующее поколение. Путем скрещивания родителей поучаются особи нового поколения, к полученным особям применяется операция мутации и цикл повторяется снова.

Цикл заканчивается, если построена особь, являющаяся решением или в случае выполнения какого-либо другого условия (количество шагов цикла, достаточная «близость» к решению и т.д.).

Для того, чтобы выделить часть особей в качестве родителей, на этапе селекции для каждой особи вычисляется фитнес-функция. После ее вычисления можно применить различные методы отбора потенциальных родителей [3].

Наиболее популярный метод – *ранговый*. Особи сортируются по значению фитнес-функции и получают номер в этом порядке. Вероятность выбора особи пропорциональна ее номеру.

Турнирный метод. Выбор каждой особи – это лучшая из выбранных случайным образом M особей. Обычно, M = 2.

Отбор усечением. Особи сортируются по

значению фитнес-функции и выбираются N заданных лучших особей. Далее из них случайным образом выбирают нужное количество.

В целом генерация поколений для всех приведенных примеров осуществляется следующим образом:

1. Начальное поколение формируется случайным образом.

2. Для каждой особи поколения считается фитнес-функция f (именно здесь алгоритмы для наших примеров слегка различаются, так как фитнес-функции для примеров различаются). Очевидно, что нам нужно добиться получения значения фитнес-функции, равной 0. Сортируем популяцию по возрастанию фитнес-функции f . В случае всех наших примеров, чем меньше значение f, тем больше шансов у особи на выживание и размножение.

3. Случайным образом выбираем двух родителей, с помощью нормального распределения. Номер родителей вычисляем по формуле:

$$i = \left| \frac{\text{random}(2*k) + \text{random}(2*k) + \text{random}(2*k) + \text{random}(2*k) + \text{random}(2*k)}{5} - k \right|,$$

где k – количество особей в популяции;

random(2*k) – случайное число от 0 до 2*k.

Среднее арифметическое случайных величин имеет нормальное распределение, пик находится в нуле, далее функция убывает.

4. Два родителя скрещиваются (например, при помощи операции кроссинговер) и образуют представителя нового поколения следующим образом:

$P_1 = a_{1,1}a_{1,2}..a_{1,i}a_{1,i+1}..a_{1,n}$ – первый родитель;

$P_2 = a_{2,1}a_{2,2}..a_{2,i}a_{2,i+1}..a_{2,n}$ – второй родитель;

$\text{Child} = a_{1,1}a_{1,2}..a_{1,i}a_{2,i+1}..a_{2,n}$ – потомок.

Здесь i определяется по формуле:

$$i = random(1..(n-1))$$

5. Часть особей для нового поколения образуется в результате мутаций. Принцип мутации следующий:

i = random (0..k) - номер мутирующей особи;

j = random (1..n) - номер мутирующей хромосомы;

b – новое случайное значение хромосомы (диапазон возможных значений зависит от особенностей задачи);

$P_i = a_{1,1}a_{1,2}..a_{1,i-1}ba_{1,i+1}..a_{1,n}$ – мутировавшая особь.

6. Формируем новое поколение как объединение особей, полученных в результате скрещивания, мутации и наиболее сильных особей предыдущего поколения (именно это обстоятельство обеспечивает нам неубывание min(f)).

Повторяем все действия, начиная с шага 2, необходимое число раз, пока не получим требуемое значение фитнес-функции.

Заключение

Окружающий мир поражает нас своей сложностью и размахом своих проявлений. Чем глубже мы исследуем самих себя и мир вокруг нас, тем более очевидной становится непостижимость окружающих нас явлений. К примерам такого рода можно отнести сложные социальные системы, нейронные и иммунные системы, разнообразные связи между видами. Многие из этих явлений можно объяснить единой теорией эволюции. Именно поэтому генетические алгоритмы продолжают оставаться популярными методами поиска решений для задач различного типа. Авторы попытались проследить соответствие между принципами теории естественного отбора Дарвина и работой генетического алгоритмов. В качестве примеров были выбраны несколько известных задач из теории чисел.

Разумеется, эволюция биологических систем не единственный «источник вдохновения» для создателей новых методов, моделирующих природные процессы. В последнее время широкое распространение получили

нейронные сети, основанные на моделировании поведения нейронов в мозге. Современные нейронные сети широко используются для ряда задач распознавания образов. Однако круг прикладных задач, решаемых с помощью современных нейронных сетей, несколько однобок, хотя абсолютно очевидно, что исследователям просто не хватает воображения, чтобы более точно понять механизм поведение нейронов в мозге и расширить сферу применения нейронных сетей для более широкого круга прикладных задач.

Авторы надеются, что заинтересовавшиеся предметом читатели найдут и другие методы, моделирующие природные процессы и попытаются использовать их для решения прикладных задач в самых разных отраслях человеческой деятельности.

Использованная литература:

1. Егоров К.В., Царев Ф.Н., Шалыто А.А. Применение генетического программирования для построения автоматов управления системами со сложным поведением на основе обучающих примеров и спецификации. // Научно-технический вестник Санкт-Петербургского государственного университета информационных технологий, механики и оптики, № 5 (69), 2010, с. 81-86.

2. Царев Ф.Н. Метод построения управляющих конечных автоматов на основе тестовых примеров с помощью генетического программирования. // Инфрмационно-управляющие системы, 2010, № 5.

3. Blickle T., Thiele L. A Comparison of Selection Schemes used in Genetic Algorithm. 1995, 2 Edition.

4. Darwin C.R. On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life. John Murray, London, 1859.

5. Fogel L.J., Owens A.J., Walsh M.J. Artificial Intelligence through Simulated Evolution. JohnWiley, NY, 1966.

6. George F. Luger. Artificial Intelligence: Structures and Strategies for Complex Problem Solving. Addison Wesley, 2004.

7. Grafeeva N., Grigorieva L., Kalinina-Shuvalova N. Genetic algorithms and genetic programming, International Journal of Advanced Computer Science, Vol. 3, No. 9, September, 2013.

8. Grafeeva N., Grigorieva L., Khristoforov V. Overview of classes of problems, effectively solved by the genetic programming. // European Applied Sciences, 2013, № 4.

9. Holland J. H. Adaptation in natural and artificial systems. University of Michigan Press, Ann Arbor, 1975.

10. Koza J. R. Genetic programming. On the Programming of Computers by Means of Natural Selection. MA.:The MIT Press. 1998.

11. Mitchell M. An Introduction to Genetic Algorithms. The MIT Press, MA. 1996.

Окружающий мир поражает нас своей сложностью и размахом своих проявлений. Чем глубже мы исследуем самих себя и мир вокруг нас, тем более очевидной становится непостижимость окружающих нас явлений. К примерам такого рода можно отнести сложные социальные системы, нейронные и иммунные системы, разнообразные связи между видами. Многие из этих явлений можно объяснить единой теорией эволюции. Именно поэтому генетические алгоритмы продолжают оставаться популярными методами поиска решений для задач различного типа. Авторы надеются, что заинтересовавшиеся предметом читатели найдут и другие методы, моделирующие природные процессы и попытаются использовать их для решения прикладных задач в самых разных отраслях человеческой деятельности.