УДК 004.021 + 519.233.5

Дивинец А.А., Кирбиц И.В., Лысюк А.Н.

ГЕНЕТИЧЕСКИЕ АЛГОРИТМЫ В ЗАДАЧЕ ОПРЕДЕЛЕНИЯ КОЭФФИЦИЕНТОВ РЕГРЕССИОННОГО УРАВНЕНИЯ

В работе рассмотрена проблема нахождения коэффициентов регрессионных уравнений, а также показан один из способов решения данной проблемы с использованием, так называемых, генетических алгоритмов, способных свести сложную математическую задачу к обычной алгоритмической процедуре и тем самым значительно упростить исходное решение задачи.

Основной целью регрессионного анализа является построение адекватной регрессионной модели[1], которая представляет собой параметрическую функцию следующего вида:

$$f\left(\mathbf{W},\mathbf{X}\right)+\varepsilon=y\,, \tag{1}$$

где **W** – вектор параметров, **X** – вектор свободных переменных, y – зависимая переменная, ε – регрессионный остаток. Из (1) следует, что для построения регрессионной модели необходимо решить три важные задачи:

1) определить вид функции f;
2) выделить набор свободных переменных, существенно влияющих на математическое ожидание зависимой величины, то есть задать вектор **X**;
3) подобрать значения параметров вектора **W**.

Первая задача решается на основании теоретических или эмпирических предпосылок. Вторая задача решается методами дисперсионного, корреляционного и другими видами анализа. Для непосредственной идентификации модели, то есть определении вектора **W**, рассматривают некоторую функцию $\phi(\mathbf{W})$, которую требуется минимизировать. В качестве примера, можно привести хорошо известную функцию суммы квадратов остатков:

$$\phi\left(\mathbf{W}\right)=\sum_{i=1}^{n}\left(y_i-f\left(\mathbf{W},\mathbf{X_i}\right)\right)^2=\sum_{i=1}^{n}\left(\varepsilon_i\right)^2,\ i=\overline{1...n}\,, \tag{2}$$

где n –общее количество наблюдений случайных величин y и **X**. В случае линейности рассматриваемых функций f(**W**, **X**) и ϕ(**W**), задача определения коэффициентов **W** имеет тривиальное решение, в отличие от нелинейных регрессионных моделей, для решения которых зачастую применяются приближенные алгоритмы, такие, как, например, метод Ньютона-Гаусса. На наш взгляд такие методы обладают тем недостатком, что имеют медленную сходимость и сложны в реализации. Зачастую использование рандомизации позволяет достичь аналогичных результатов с гораздо меньшими затратами.

Поэтому, для решения поставленной задачи, то есть определения значений коэффициентов регрессии используем генетические алгоритмы, которые, как известно, являются вероятностными и обладают хорошей сходимостью при решении задач оптимизации – сведению целевых функций к минимуму или максимуму[2].

При использовании генетических алгоритмов, исходная задача формализуется таким образом, чтобы её решение могло быть закодировано в виде вектора *генов* (*генотипа*), где каждый ген, как правило, представляет собой отдельный бит. Некоторым, обычно случайным, образом создаётся множество генотипов начальной популяции, которая оценивается с использованием *функции приспособленности*, в результате чего с каждым генотипом ассоциируется определённое значение приспособленности, которое

определяет то, насколько хорошо решается поставленная задача с использованием соответствующего генотипа.

Из полученного множества генотипов выбираются наиболее приспособленные экземпляры или особи, к которым применяются генетические операторы, такие как скрещивание и мутация, результатом чего является получение нового набора генотипов. Далее весь процесс повторяется до тех пор, пока не будет найдена особь с требуемым по условию задачи уровнем приспособленности.

Рассмотрим некоторые особенности использования генетических алгоритмов в задаче нахождения параметров регрессионных моделей. В качестве целевой функции приспособленности будем рассматривать функцию $\phi(\mathbf{W})$. Генотипы будут состоять из n групп генов, где каждая отдельная группа будет представлять собой координату вектора $\mathbf{W}$. При этом все генетические операции будут выполняться с генами только внутри отдельных групп. Можно привести аналогию с биологическими организмами, где гены, отвечающие за цвет глаз, не скрещиваются с генами, отвечающими за размер ушей. Зная, как интерпретировать значения генов для некоторого генотипа и особенности выполнения генетических операций, перейдем к непосредственному описанию функционирования алгоритма.

Шаг 1. Инициировать начальный момент времени t=0. Случайным образом сформировать начальную популяцию, состоящую из m особей.

Шаг 2. Вычислить приспособленность каждой особи на основании соответствующего полученного значения целевой функции. Чем она меньше, тем более приспособлена особь. Произвести отсеивание k самых неприспособленных особей.

Шаг 3. Из оставшейся популяции случайным образом выбрать две особи и выполнить над ними операции кроссовера и мутации.

Шаг 4. Поместить одну из результирующих особей в новую популяцию.

Шаг 5. Выполнить операции, начиная с шага 3, k раз;

Шаг 6. Увеличить номер текущей эпохи t=t+1;

Шаг 7. При достижении функцией приспособленности минимального значения завершить работу, иначе перейти к шагу 2.

В ходе испытания алгоритма, полученные результаты совпали с теоретически заданными параметрами модели, что, несомненно, подтверждает корректность работы данного алгоритма. Кроме того было выявлено, что скорость сходимости алгоритма главным образом зависит от размеров исходной популяции.

Список литературы

1. Соколов, Г.А. Введение в регрессионный анализ и планирование регрессионных экспериментов в экономике: учебное пособие / Г.А. Соколов, Р.В. Сагитов. – М.: Инфра-М, 2010. – 208 с.
2. Гладков, Л.А. Генетические алгоритмы / Л.А. Гладков, В.В. Курейчик, В.М. Курейчик. – М.: ФИЗМАТЛИТ, 2010. – 368 с.

Дивинец А.А., студент факультета электронно-информационных систем БГТУ, Брест, Беларусь, e-mail: alexdiv_91@mail.ru.

Кирбиц И.В., студентка факультета электронно-информационных систем БГТУ, Брест, Беларусь, e-mail: alexdiv@pisem.net.

Лысюк А.Н., ассистент кафедры «ЭВМ и Системы» БГТУ, Брест, Беларусь, e-mail: lysiuk.andrei@gmail.com.