УДК 336.22:519.68:332.12

С. Г. Зеленская,
Б. Г. Преображенский,
А. А. Спирячин

ОПТИМИЗАЦИЯ ВЛИЯНИЯ ФАКТОРОВ НА ПРОЦЕСС НАЛОГОВЫХ ПОСТУПЛЕНИЙ В БЮДЖЕТ РЕГИОНА С ИСПОЛЬЗОВАНИЕМ НЕЙРОННОЙ СЕТИ И ГЕНЕТИЧЕСКОГО АЛГОРИТМА

В статье предложен метод, который позволяет оптимизировать влияние факторов налогооблагаемой базы на формирование бюджета региона с использованием возможностей нейронной сети и генетического алгоритма.

Ключевые слова: *налоговая база, оптимизация, искусственная нейронная сеть, генетический алгоритм*

UDK 336.22:519.68:332.12

S. G. Zelenskay,
B. G. Preobrazhensky,
A. A. Spirychin

OPTIMIZATION OF THE INFLUENCE OF THE PROCESS OF TAX REVENUES OF THE REGION WITH THE USE OF ARTIFICIAL NEURAL NETWORK AND GENETIC ALGORITHM

The article suggests a method, which allows to optimize the impact of the taxable base of the region with use of opportunities of the neural network and genetic algorithm.

Key words: *tax base, optimization, artificial neural network, genetic algorithm.*

Оптимизация налоговой базы является доминантным условием повышения эффективности налоговой системы региона. Формальное описание налоговой системы в виде математической модели процесс чрезвычайно сложный. Однако известен ряд факторов влияющих на эффективность наполнения налогами и сборами бюджетной системы регионов. Метод оптимизации, предложенный здесь, состоит из двух этапов: первый — использование нейронной сети для выражения математической модели налоговой системы[1], второй — оптимизация методом эволюционного поиска глобального экстремума значений, являющихся результатом воздействия количественных показателей факторов, влияющих на математическую модель налоговой системы, которая будет представлена обученной нейронной сетью.

Первый этап состоит в построении искусственной нейронной сети и её обучении.

Искусственные нейронные сети (ИНС) — математические модели, а также их программные или аппаратные реализации, построенные по принципу организации и функционирования биологических нейронных сетей — сетей нервных клеток живого организма.

С точки зрения машинного обучения, нейронная сеть представляет собой частный случай методов распознавания образов, дискриминантного анализа, методов кластеризации и т. п. С математической точки зрения, обучение нейронных сетей — это многопараметрическая задача нелинейной оптимизации. С точки зрения кибернетики, нейронная сеть используется в задачах адаптивного управления и как алгоритмы для робототехники. С точки зрения развития вычислительной техники и программирования, нейронная сеть — способ решения проблемы эффективного параллелизма. А с точки зрения искусственного интеллекта, ИНС является основой философского течения коннективизма и основным направлением в структурном подходе к изучению возможности построения (моделирования) естественного интеллекта с помощью компьютерных алгоритмов.

Нейронные сети не программируются в привычном смысле этого слова, они обучаются. Возможность обучения — одно из главных преимуществ нейронных сетей перед традиционными алгоритмами. Технически обучение заключается в нахождении коэффициентов связей между нейронами. В процессе обучения нейронная сеть способна выявлять сложные зависимости между входными и выходными данными, а также выполнять обобщение. Это значит, что в случае успешного обучения сеть сможет вернуть верный результат на основе данных, которые отсутствовали в обучающей выборке, а также неполных и/или «зашумленных», частично искаженных данных.

Способности к обобщению и выделению скрытых зависимостей между входными и выходными данными нейронной сети позволяют создать нейронную сеть, представляющую с некоторой точностью математическую модель налоговой базы региона. Следует отметить, что нейронная сеть должна быть обучена на статистических данных, определяющих налоговую систему региона. К таким данным могут быть отнесены количественные показатели факторов, составляющих налогооблагаемую базу региона, и количественный показатель поступивших в бюджет налогов за несколько лет. Сложность описания математической модели налоговой системы в том числе определяется наличием такого фактора, как уклонение от уплаты налогов. Конечно, политика, направленная на ужесточение законодательства в этой сфере, может принести некоторые плоды, однако через определённое время окажется, что законы можно обойти. Также следует учитывать затраты на раскрытие преступлений, связанных с уклонением от уплаты налогов. Эти затраты тоже являются фактором, влияющим на сбор налогов.

В нашем случае нейронная сеть способна выявить закономерности изменения налоговой базы, однако оптимизацию разумнее осуществлять на краткосрочный период, поскольку, несмотря на инертность, налоговая система претерпевает множество изменений с течением времени. Это связано в том числе и с воздействием внешних и случайных факторов, с развитием экономики, повышением правовой культуры граждан.

Структура нейронной сети определяется конечной целью исследования — обобщить влияние факторов и представить математическую модель налоговой системы как «чёрный ящик». Налоговые поступления в бюджет являются количественным показателем. Целесообразно использовать единственный выход у нейронной сети, который будет определять объем налоговых и неналоговых поступлений в бюджетную систему на территории региона. На входы подаются количественные значения параметров налоговой системы и контингентов налогооблагаемой базы за предыдущие периоды по порядку, а на выход — совокупные объемы налоговых и неналоговых поступлений в бюджетную систему на территории региона за эти же периоды. Для эффективного обучения сети нужно иметь достаточное количество данных. От количества входов зависит требуемое количество значений в обучающей выборке. Чтобы добиться требуемой точности при методе обратного распространения ошибки, вероятно, придётся провести несколько итераций обучения — эпох. Количество эпох будет зависеть от требуемой точности и параметров настройки нейронной сети — скорости обучения, количества скрытых слоёв, количества нейронов в скрытом слое. После того, как требуемая точность будет достигнута, проводят проверку адекватности обучения на основе тестирующей выборки. В данной выборке также присутствует пара векторов значений — входов и выходов, однако, в отличие от обучающей выборки, выходное значение в этой паре служит лишь для оценки обучения. На вход подаются лишь первые значения пары в тестирующей выборке, а по вторым проверяется адекватность обучения. После того, как тестирование закончено, делается вывод об адекватности обучения. Если обучение прошло нормально, то приступают ко второму этапу — оптимизации с помощью генетических алгоритмов, иначе — обучают сеть снова теми же данными или ищут возможность увеличить количество элементов в обучающей выборке, настраивают сеть, подбирают оптимальное число входов. Значения параметров настройки сети подбираются эмпирическим путём, поэтому он может быть довольно трудоёмким процессом [2].

После того, как сделан вывод, что сеть обучена адекватно, можно приступать ко второму этапу.

Второй этап подразумевает использование генетического алгоритма (ГА) для оптимизации влияния экзогенных и эндогенных факторов на поступление налоговых и неналоговых платежей в бюджетную систему на территории региона.

Генетический алгоритм — это одно из направлений исследований в области искусственного интеллекта, занимающееся созданием упрощенных моделей эволюции живых организмов для решения задач оптимизации. Является разновидностью эволюционных вычислений, с помощью которых решаются оптимизационные задачи с использованием методов естественной эволюции, таких как наследование, мутации, отбор и кроссинговер. Первые работы по симуляции эволюции были проведены в 1954 году Нильсом Баричелли на компьютере, установленном в Институте Продвинутых Исследований Принстонского университета [3,4]. Его работа, опубликованная в том же году, привлекла широкое внимание общественности. Сегодня становятся более популярными методы решения задач, основанные на совместном использовании нейронных сетей и генетических алгоритмов.

Задача формализуется таким образом, чтобы её решение могло быть закодировано в виде вектора («генотипа») генов, где каждый ген может быть битом, числом или неким другим объектом. В классических реализациях ГА предполагается, что генотип имеет фиксированную длину. Однако существуют вариации ГА, свободные от этого ограничения. В нашем случае генотип — это выраженное в двоичном виде количественное значение факторов, определяющих налоговую базу региона. Следует отметить, что и как в любой задаче оптимизации должны быть введены ограничения на значения факторов. В краткосрочный период оценить границу разброса их значений не составит труда.

Некоторым, обычно случайным, образом создаётся множество генотипов начальной популяции — случайное значение факторов из области определения. Они оцениваются с использованием «функции приспособленности», в результате чего с каждым генотипом ассоциируется определённое значение («приспособленность»), которое определяет, насколько хорошо фенотип, им описываемый, решает поставленную задачу. В данном случае значение функции приспособленности, а именно совокупное значение поступлений средств в бюджет от налогов, будет вычислять обученная нейронная сеть.

Из полученного множества решений («поколения») с учётом значения «приспособленности» выбираются решения (обычно лучшие особи имеют большую вероятность быть выбранными), к которым применяются «генетические операторы» (в большинстве случаев «скрещивание» — crossover и «мутация» — mutation), результатом чего является получение новых решений. Для них также вычисляется значение приспособленности, затем производится отбор («селекция») лучших решений в следующее поколение. Следует отметить, что мутация необходима для разнообразия генотипа, а также она может коррелировать с вероятностями воздействия случайных факторов на значение налоговой базы.

Классический генетический алгоритм (также называемый элементарным или простым генетическим алгоритмом) состоит из следующих шагов:

1) инициализация, или выбор исходной популяции хромосом;

2) оценка приспособленности хромосом в популяции — расчет функции приспособленности для каждой хромосомы;

3) проверка условия остановки алгоритма;

4) селекция хромосом — выбор тех хромосом, которые будут участвовать в создании потомков для следующей популяции;

5) применение генетических операторов — мутации и скрещивания;

6) формирование новой популяции;

7) выбор «наилучшей» хромосомы.

Блок-схема основного генетического алгоритма изображена на рисунке.

Как видно из рисунка, шаги 2–6 повторяются циклически.

Генетические алгоритмы служат, главным образом, для поиска решений в многомерных пространствах поиска, каким и является обученная нейронная сеть, выражающая математическую модель налоговой базы.

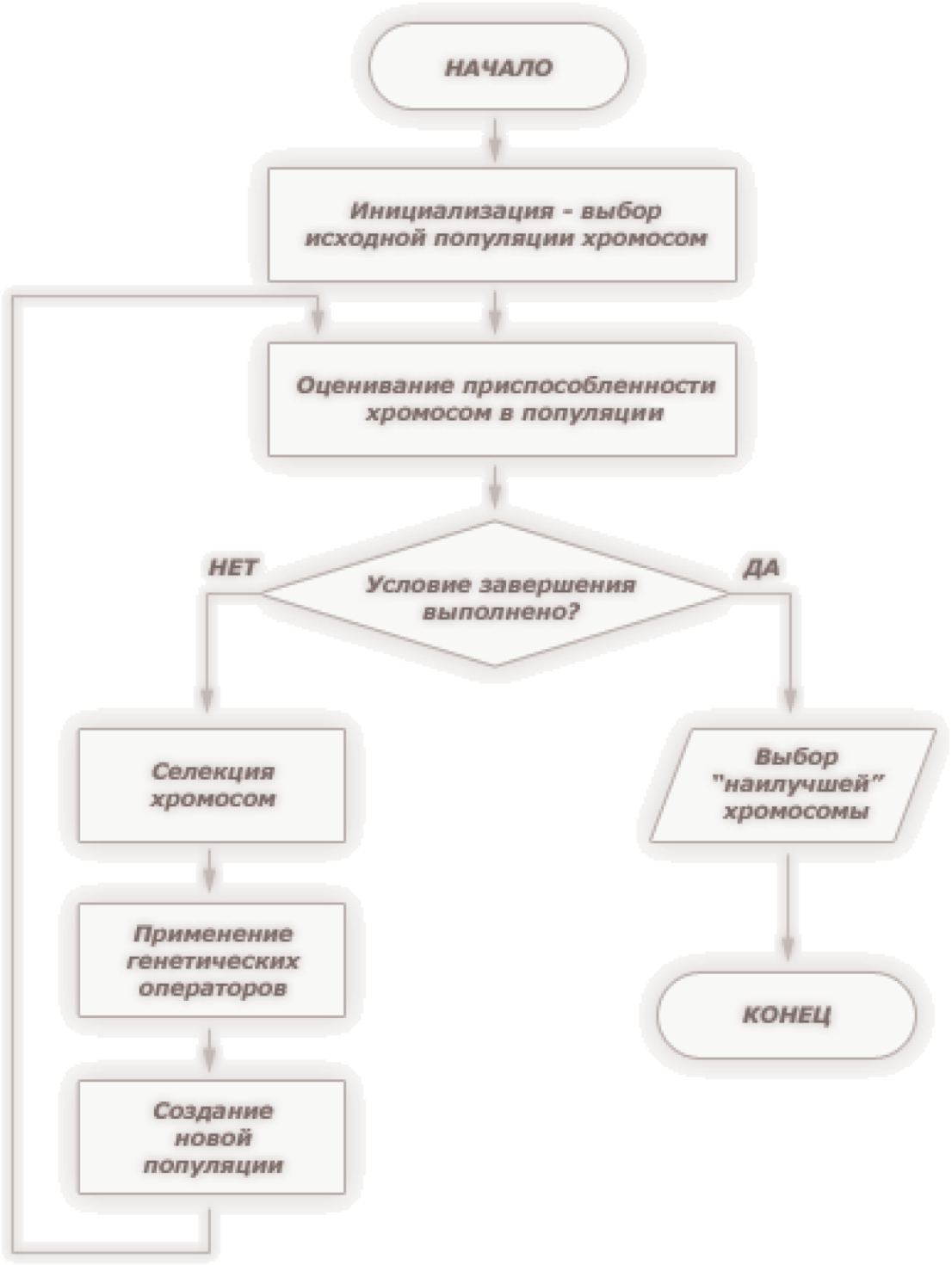


Рис. Блок-схема генетического алгоритма

Операция скрещивания заключается в обмене фрагментами цепочек между двумя родительскими хромосомами. Пары родителей для скрещивания выбираются из родительского пула случайным образом так, чтобы вероятность выбора конкретной хромосомы для скрещивания была равна вероятности p_c. Например, если в качестве родителей случайным образом выбираются две хромосомы из родительской популяции численностью N, то $p_c = 2/N$. Аналогично, если из родительской популяции численностью N выбирается 2z хромосом ($z < N/2$), которые образуют z пар родителей, то $p_c = 2z/N$. Обратим внимание, что если все хромосомы текущей популяции объединены в пары до скрещивания, то $p_c = 1$. После опе-

рации скрещивания родители в родительской популяции замещаются их потомками.

Операция мутации изменяет значения генов в хромосомах с заданной вероятностью p_m. Это приводит к инвертированию значений отобранных генов с 0 на 1 и обратно. Значение p_m, как правило, очень мало, поэтому мутации подвергается лишь небольшое количество генов. Скрещивание — это ключевой оператор генетических алгоритмов, определяющий их возможности и эффективность. Мутация играет более ограниченную роль. Она вводит в популяцию некоторое разнообразие и предупреждает потери, которые могли бы произойти вследствие исключения какого-нибудь значимого гена в результате скрещивания.

Селекция – это выбор тех хромосом, которые будут участвовать в создании потомков для следующей популяции, т.е. для очередного поколения. Такой выбор производится согласно принципу естественного отбора, по которому наибольшие шансы на участие в создании новых особей имеют хромосомы с наибольшими значениями функции приспособленности. Существуют различные методы селекции, рассмотрим их.

Основанный на принципе колеса рулетки метод селекции считается для генетических алгоритмов основным методом отбора особей для родительской популяции с целью последующего их преобразования генетическими операторами, такими как скрещивание и мутация. Несмотря на случайный характер процедуры селекции, родительские особи выбираются пропорционально значениям их функций приспособленности: каждой хромосоме сопоставлен сектор колеса рулетки, величина которого устанавливается пропорциональной значению функции приспособленности данной хромосомы, поэтому, чем больше значение функции приспособленности, тем больше сектор на колесе рулетки. Отсюда вытекает, что чем больше сектор на колесе рулетки, тем выше шанс, что будет выбрана именно эта хромосома.

Вероятность выбора каждой особи равна, таким образом,

$$p_s = m/M,$$

где m — значение хромосомы особи; M — сумма значений всех хромосом особей в популяции.

Слабая сторона этого метода заключается в том, что особи с очень малым значением функции приспособленности слишком быстро исключаются из популяции, что может привести к преждевременной сходимости генетического алгоритма. В связи с вышесказанным, созданы и используются альтернативные алгоритмы селекции.

При турнирной селекции все особи популяции разбиваются на подгруппы с последующим выбором в каждой из них особи с наилучшей приспособленностью. Различаются два способа такого выбора: детерминированный выбор и случайный выбор. Детерминированный выбор осуществляется с вероятностью равной 1, а случайный выбор — с вероятностью меньшей 1. Подгруппы могут иметь произвольный размер, но чаще всего популяция разделяется на подгруппы по 2–3 особи в каждой.

Турнирный метод пригоден для решения задач как максимизации, так и минимизации функции. Помимо того, он может быть легко распространен на задачи, связанные с многокритериальной оптимизацией, т.е. на случай одновременной оптимизации нескольких функций. В турнирном методе допускается изменение размера подгрупп, на которые подразделяется популяция. Исследования подтверждают, что турнирный метод действует эффективнее, чем метод рулетки.

При ранговой селекции особи популяции ранжируются по значениям их функции приспособленности. Это можно представить себе как отсортированный список особей, упорядоченных по направлению от наиболее приспособленных к наименее приспособленным (или наоборот), в котором каждой особи приписывается число, определяющее ее место в списке и называемое рангом. Количество копий каждой особи, введенных в родительскую популяцию, рассчитывается по априорно заданной функции в зависимости от ранга особи.

Элитарная стратегия заключается в защите наилучших хромосом на последующих итерациях. В классическом генетическом алгоритме самые приспособленные особи не всегда переходят в следующее поколение. Это означает, что новая популяция не всегда содержит хромосому с наибольшим значением функции приспособленности из предыдущей популяции. Элитарная

стратегия применяется для предотвращения потери такой особи. Эта особь гарантированно включается в новую популяцию.

Генетический алгоритм с частичной заменой популяции, иначе называемый генетическим алгоритмом с зафиксированным состоянием, характеризуется тем, что часть популяции переходит в следующее поколение без каких-либо изменений. Это означает, что входящие в эту часть хромосомы не подвергаются операциям скрещивания и мутации. Часто в конкретных реализациях алгоритма данного типа на каждой итерации заменяются только одна или две особи вместо скрещивания и мутации в масштабе всей популяции[5].

Следует отметить, что возможно совмещение методов отбора для достижения наилучшего результата применимо к данной задаче.

Итак, нами рассмотрены два этапа оптимизации влияния факторов на налоговую базу региона. Первый этап состоял в описании обучения нейронной сети для получения неявной математической модели как «чёрного ящика». Второй — в описании возможностей использования эволюционного генетического алгоритма для оптимизации значений факторов, поступающих на входы «чёрного ящика». Функция приспособленности — это по сути структура нейронной сети, а значение выхода нейронной сети соответствует значению этой функции приспособленности. Использование двух методов искусственного интеллекта позволит при достоверности данных для обучения и относительной стабильности системы налогообложения оптимизировать влияние случайных, экзогенных, эндогенных и структурных факторов региональной налоговой базы на налоговые и неналоговые поступления в бюджетную систему региона.

ЛИТЕРАТУРА

1. Зеленская С. Г., Преображенский Б. Г., Спирячин А. А. Использование математической модели искусственной нейронной сети для текущего прогнозирования налоговой базы региона// Регион: системы, экономика, управление –Воронеж: ИПЦ «Научная книга», 2012, №1(16). –С. 126–131.

2. http://ииклуб.рф/neur-2.html — Нейронные сети: алгоритм обратного распространения/ С. Короткий.

3. Barricelli, Nils Aall (1954). «Esempi numerici di processi di evoluzione». Methodos: 45–68.

4. Barricelli, Nils Aall (1957). «Symbiogenetic evolution processes realized by artificial methods». Methodos: 143–182.

5. http://www.AIPortal.ru — Портал искусственного интеллекта.

6. Искусственные нейронные сети. Теория и практика/ Круглов В. В., Борисов В. В. — М.: Горячая линия — Телеком, 2001. —382 с.

7. Мак-Каллок У. С., Питтс В., Логическое исчисление идей, относящихся к нервной активности // В сб.: «Автоматы» под ред. К. Э. Шеннона и Дж. Маккарти. — М.: Издво иностр. лит., 1956. — С.363–384 (Перевод английской статьи 1943 г.).

8. Петров А. П. О возможностях перцептрона // Известия АН СССР, Техническая кибернетика. — 1964. — № 6. — С. 114–121.