

А.А. Левтеров, к.т.н., с.н.с., вед. научн. сотр., НУГЗУ

ПРОГНОЗИРОВАНИЕ ЧРЕЗВЫЧАЙНОЙ СИТУАЦИИ ПРИРОДНОГО ХАРАКТЕРА С ИСПОЛЬЗОВАНИЕМ ГЕНЕТИЧЕСКИХ АЛГОРИТМОВ

(представлено д-ром техн. наук Прохачем Э.Е.)

Предложено использование генетических алгоритмов при прогнозировании факторов ЧС. Рассмотрена методика расчета краткосрочного прогноза площади, ограниченной кромкой пожара с использованием генетического алгоритма.

Ключевые слова: генетический алгоритм, краткосрочный прогноз, чрезвычайная ситуация.

Постановка проблемы. Использование традиционных методов статистики при прогнозировании ЧС [10, 12], а также математических моделей предложенных в [13] подразумевают ряд ограничений и определенные требования к целевым функциям, а также невозможность повышения точности прогноза при изменении параметров, например, в процессе развития ЧС. Данные ограничения, при поиске оптимальных решений, не позволяют повысить точность прогноза до требуемого значения. При применении генетических алгоритмов (ГА), основанных на механизмах естественного отбора и наследования, позволяет избежать ряд ограничений, и тем самым повысить точность прогноза.

В ГА используется эволюционный принцип [1], где поиск экстремума целевой функции осуществляется одновременно по многим направлениям путем использования популяции возможных решений, а переход от одной популяции к другой позволяет избежать попадания в локальный оптимум, при этом ГА характеризуется полиномиальной сложностью вычислений.

Применение ГА решает проблему, используя процесс, подобный биологическому развитию. Он работает как рекомбинация и мутация генетических последовательностей. Рекомбинация и мутация – генетические операторы, т.е. они управляют генами (последовательность кодов), содержащими всю информацию, необходимую для того, чтобы создать функциональный организм с определенными характеристиками (генотипом) [1].

В случае генетической оптимизацией, используемой для решения задач, связанных с прогнозированием, последовательность кодов обычно принимает форму ряда чисел. Как и в процессе биологического отбора (где менее пригодные члены популяции оставляют меньше потомства), менее пригодные решения удаляются. При этом более при-

годные решения размножаются, создавая другое поколение решений, которое может содержать несколько лучшие решения, чем предыдущие. Процесс рекомбинации, случайной мутации и отбора является чрезвычайно действенным механизмом решения задачи. Применение генетических алгоритмов базируется на следующих принципах:

- исследуемый процесс, зависит от множества факторов, и вывести однозначную его формулу невозможно;

- не существует единой закономерности, которая действовала бы на всём протяжении процесса;

- на небольших участках времени могут наблюдаться закономерности, которые сложно формализовать;

- выдвигаются некоторые простые правила, по которым можно сделать прогноз;

- правила собираются в группы, образуя так называемые «хромосомы», по аналогии с биологией;

- каждая из хромосом тестируется на имеющейся базе событий. Из получившихся хромосом половину (что показали самые неудачные результаты) выбрасывают;

- к удачным хромосомам применяют методики ГА – рекомбинацию и мутацию. Мутация означает случайное изменение какого-то правила, рекомбинацию – скрещивание двух хромосом, которые «обмениваются» своими частями. Таким образом создают ещё столько же новых хромосом, сколько было и родительских;

- повторяют 2 последних пункта до тех пор, пока не останутся только лучшие хромосомы [1, 2].

В идеале, по завершению всей процедуры, получим несколько стратегий, которые будут эффективными, например, при ликвидации ЧС.

Цель статьи исследование возможности применения ГА к решению задачи краткосрочного прогнозирования значений разных факторов, связанных с развитием ЧС, при минимальных временных затратах. За основу возьмем подход предложенный в [3].

Анализ последних исследований и публикаций. Исходя их данных приведенных в [11-13] и в [6-9] можно сделать вывод, что ГА при прогнозировании развития ЧС не применяются. Как показывает детальный анализ приведенных методов и способов статистического прогнозирования применение ГА в данном аспекте не противоречит логике и математике, заложенной в этих методах.

В связи с этим рассмотрим модель прогнозирования развития ЧС с применением ГА.

Постановка задачи и ее решение. Модель прогнозирования строится на основе накопленных данных двух факторов: главного d_1, d_2, \dots, d_n и вспомогательного e_1, e_2, \dots, e_n , где n – длина актуальной части ряда (количество наблюдений временного ряда – значения факторов

$$(X_{i-k}, Y_{i-k}), \dots, (X_{i-2}, Y_{i-2}), (X_{i-1}, Y_{i-1}) \rightarrow X_i \quad (5)$$

которые непосредственно используются для выполнения прогноза [3, 4].

В качестве критерия оптимальности целесообразно использовать минимальное значение средней относительной ошибки прогнозирования AFER (Average Forecasting Error Rate) [3]

$$AFER = \frac{\sum_{i=k+1}^n (d_i^* - d_i) / d_i}{n - k} \cdot 100\%, \quad (6)$$

где d_i^* d_i – предсказанное и реальное значения для i -го периода прогноза за $i = k + 1, n$.

Прогнозируемое значение для i -го периода прогноза вычисляется по формуле (4)

$$d_i^* = \frac{w_{j1} \cdot t_{j1} + w_{j2} \cdot t_{j2} + \dots + w_{jp} \cdot t_{jp}}{w_{j1} + w_{j2} + \dots + w_{jp}} \quad (7)$$

Здесь w_j – коэффициенты, учитывающие повторы в нечетких логических зависимостях, а переменные t_j определяются по формуле

$$t_i = \begin{cases} (\tilde{u}_1 + 0.5 \cdot \tilde{u}_2) / (1 + 0.5), j=1 \\ (0.5 \cdot \tilde{u}_{j-1} + \tilde{u}_j + 0.5 \cdot \tilde{u}_{j+1}) / (0.5 + 1 + 0.5), 2 \leq j \leq n - 1 \\ (0.5 \cdot \tilde{u}_{n-1} + \tilde{u}_n) / (0.5 + 1), j = n \end{cases} \quad (8)$$

где \tilde{u}_{j-1} , \tilde{u}_j , \tilde{u}_{j+1} – средние точки интервалов u_{j-1} , u_j , u_{j+1} соответственно [2, 3].

При использовании изложенного подхода особое внимание должно уделяться выбору свободных параметров модели прогнозирования, к которым относятся ее порядок k , числа коррекции диапазонов D_1 , D_2 , E_1 , E_2 , а также количества интервалов p и q , на которые разбиваются универсумы U и V соответственно.

Оптимальные значения этих параметров обеспечивают минимум показателя AFER, причем в базовой модели (6) они выбираются вручную. Для сокращения времени поиска и повышения точности результатов предлагается использовать классический ГА [1], где структура хромосомы будет иметь вид

$$S = (D_1, D_2, E_1, E_2, p, q, k), \quad (9)$$

Если для некоторой хромосомы S_i все правые части групп нечетких логических зависимостей определены (не пусты) [3], то значение функции приспособленности вычисляется как $J(S_i) = AFER$. В том случае, когда не определена правая часть хотя бы одной такой зависимости, невозможно гарантировать вычисление прогнозируемого значения для нового члена временного ряда. Такую хромосому необходимо исключить из популяции как наихудшую, вычислив ее функцию приспособленности $J(S_i)$ как сумму средней относительной ошибки прогноза $AFER$ и числа 100. Окончательно функция приспособленности принимает вид [4]

$$J(S_i) = \begin{cases} AFER, & \text{если определены все правые части} \\ & \text{логических зависимостей;} \\ AFER + 100 & \text{– в противном случае} \end{cases} \quad (10)$$

Хромосома, для которой в результате выполнения заданного числа итераций ГА достигается минимум функции приспособленности (5), определяет оптимальный набор параметров модели прогнозирования.

Следует заметить, что поиск минимального значения функции (5) классическими методами численной оптимизации является весьма затруднительным, так как область допустимых решений задачи определяется неизвестными заранее пределами изменения параметров модели прогнозирования. Кроме того, указанная область будет невыпуклой и несвязной, поскольку переменные D_1, D_2, E_1, E_2 являются непрерывными, а переменные p, q, k – дискретными (целыми).

Реализация ГА выполняется по классической схеме [2] и включает выполнение следующих шагов.

Шаг 1. Создается начальная популяция размера N из случайным образом выбранных хромосом $S_i = (i = \overline{1, N})$ вида (9).

Генерация выполняется случайным выбором аллелей для каждого гена. Если первоначальная популяция окажется неконкурентоспособной, то ГА переведет её в жизнеспособную популяцию. Для этого достаточно, чтобы значения генов находились в таких пределах, которые позволяют вычислить функцию приспособленности (10). Поэтому параметры D_1, D_2, E_1, E_2 представляются действительными числами, а количество интервалов p и q находятся в пределах от 2 до n . Порядок модели ограничен значениями от 1 до 5, что объясняется достаточно короткой длиной анализируемых временных рядов.

Шаг 2. При $g < G$, где g и G – текущее и максимальное количество генераций, вычисляется значение функции приспособленности (10) для каждой хромосомы. Затем создаются пары хромосом-родителей и осуществляется переход к следующему шагу 3. После выполнения G генераций осуществляется переход к шагу 5.

Выбор родителей основан на комбинации методов рулетки и ранговой селекции [1] и имитирует естественный отбор: в родительскую пару включаются хромосомы с лучшими значениями функции приспособленности $J(S_i)$. Вероятность P_i выбора родителя с хромосомой S_i можно рассчитать по формуле

$$P_i = [J_{\max} - J(S_i)] / \sum_{j=1}^N [J_{\max} - J(S_j)] \quad (11)$$

где J_{\max} – наихудшее значение функции $J(S_i)$ среди хромосом текущего поколения. Тогда выбор родителя будет состоять в определении хромосомы с меньшим значением функции (10) из двух случайно отобранных с учетом вероятностей P_i ($i = \overline{1, N}$). Пара найденных таким образом хромосом-родителей будет использоваться для скрещивания.

Шаг 3. Выполняются операции скрещивания и мутации для созданных пар хромосом-родителей текущей популяции.

Скрещивание заключается в передаче участков генов от родителей к потомкам. Для этого формируется случайное число N_c , равномерно распределенное на отрезке $[0, 1]$, которое сравнивается с заданным коэффициентом скрещивания R_c . Если $R_c > N_c$, то операция скрещивания выполняется в точке z (номер гена в хромосоме), также выбранной случайным образом.

Мутация выполняется с некоторой вероятностью P_m , при которой происходит замена аллеля случайным значением. Оно выбирается с равной вероятностью в области определения гена. Заметим, что именно благодаря мутации расширяется область генетического поиска.

При выполнении мутации задается коэффициент мутации R_m и генерируется случайное число N_m , равномерно распределенное на отрезке $[0, 1]$. Если $R_m > N_m$, то случайным образом выбирается точка мутации z .

Шаг 4. Создается новая популяция, дополненная хромосомами-потомками. Затем хромосомы с худшими значениями функции приспособленности (5) отбрасываются. В итоге размер популяции сокращается до первоначального.

Шаг 5. Из полученной популяции выбирается хромосома с минимальным значением функции приспособленности (5).

Алгоритм заканчивает работу после выполнения заданного количества генераций G , причем можно показать, что вычислительная

сложность предлагаемого генетического алгоритма пропорциональной величине $O(G^2N + GN^2)$, т.е. является полиномиальной и наиболее заметно зависит от размера популяции N .

Эффективность применения ГА, для прогнозирования развития ЧС можно показать на следующем примере. В качестве прогнозируемого параметра выступает площадь ограниченная кромкой пожара, которая меняется со временем. Реальные значения получены на основании данных оперативно-диспетчерской службы оперативно-координационного центра ГСЧС.

Модель прогнозирования строится по данным площади зоны ЧС за отрезок времени равный 24 часам.

В качестве вспомогательного фактора рассматривается скорость ветра. Для случая $N=30$ и $G=1000$ получены следующие оптимальные параметры модели:

$D_1 = -1,138$; $D_2 = 0,485$; $E_1 = -0,918$; $E_2 = 1,188$; $p = 14$; $q = 14$; $k = 2$. Средняя относительная ошибка прогнозирования по известным значениям ряда составляет $AFER = 2,169\%$. Прогнозируемое и реальное значение площади зоны ЧС через 24 часа, от момента начала измерений равны 15,5 и 15,4 га соответственно. Относительная ошибка прогноза составляет 1,283%. Графическая зависимость результатов прогнозирования по обучающей выборке приведена на рисунке.

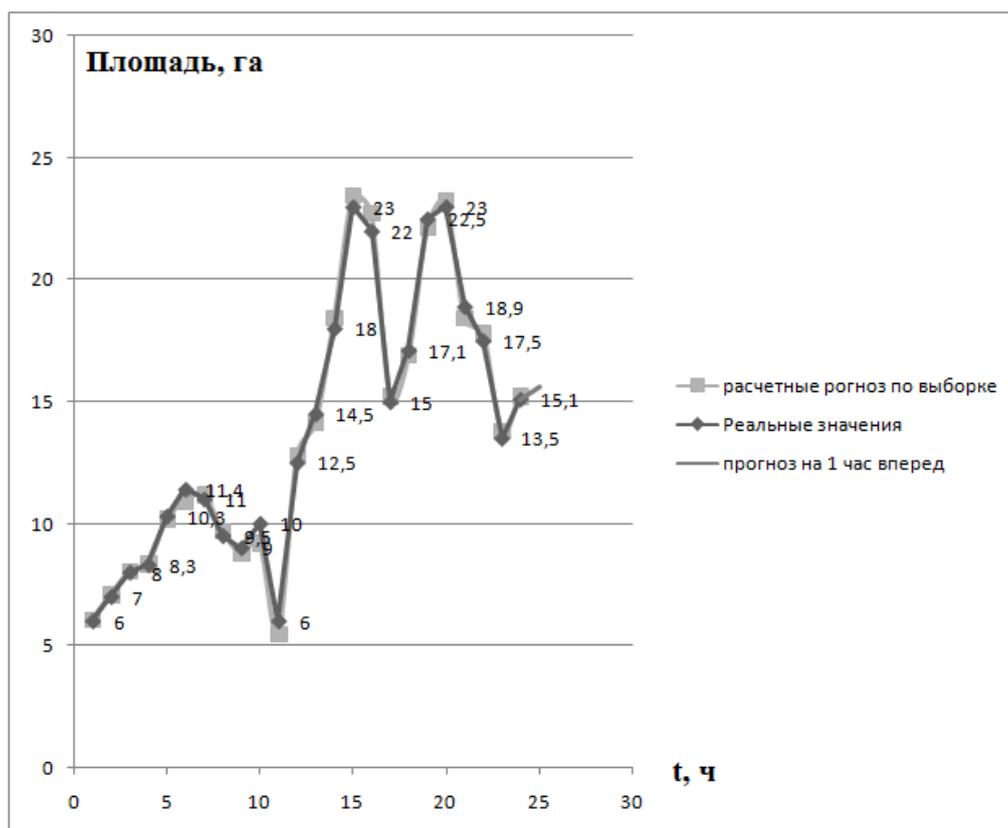


Рис. 1. Графическая зависимость реальных и прогнозируемых значений

Стоит отметить, что изменение найденных чисел корректировки границ диапазонов в небольших пределах приводит к увеличению ошибки AFER.

Таким образом, применение ГА поиска оптимальных параметров двухфакторной модели прогнозирования на основе аппарата теории нечетких множеств позволяет повысить качество прогноза за счет минимизации функции приспособленности – средней относительной ошибки прогнозирования AFER.

Выводы. Применение методики расчета с использованием ГА, построенного на нечетких множествах, при прогнозировании факторов ЧС, позволяет в некоторых случаях, особенно при краткосрочном прогнозировании повысить точность прогноза, о чем свидетельствует значение AFER.

ЛИТЕРАТУРА

1. Нейронные сети, генетические алгоритмы и нечеткие системы / Рутковская Д., Пилиньский М., Рутковский Л. – М.: Горячая линия – Телеком, 2006. – 452 с.
2. Lee L.W. Handling forecasting problems based on two-factors high-order fuzzy time series / Wang L. H., Chen S. M. // IEEE Transactions on fuzzy Systems. 2006, V.14. – № 3. – P. 468-477.
3. Демидова Л.А. Разработка двухфакторной модели прогнозирования временных рядов с использованием генетического алгоритма / Л.А. Демидова, Т.С. Скворцова. – Математическое и программное обеспечение вычислительных систем: межвуз. сб. М.: Горячая линия – Телеком, 2008, С. 99-108.
4. Fraser A.S. Simulation of genetic systems / Fraser A.S. – J. of Theor. Biol. 1962. – Vol. 2. – P. 29-46.
5. Заде Л. Понятие лингвистической переменной и его применение для принятия приближенных решений / Заде Л. – М.: Мир, 1976. – 165 с.
6. Батищев Д.И. Генетические алгоритмы решения экстремальных задач / Батищев Д.И. – Воронеж, 1995. – 65 с.
7. Рутковская Д. Нейронные сети, генетические алгоритмы и нечеткие системы / Рутковская Д., Пилиньский М., Рутковский Л. – М.: Горячая линия. – Телеком, 2004. – 452 с.
8. Лукашин Ю. П. Адаптивные методы краткосрочного прогнозирования временных рядов: Учеб. Пособие / Ю.П. Лукашин М.: Финансы и статистика, 2003. – 416 с.
9. Алтунин А.Е. Модели и алгоритмы принятия решений в нечетких условиях / А.Е. Алтунин, М.В. Семухин: Монография. – Тюмень: Издательство Тюменского государственного университета, 2000. – 352 с.

10. Белов П.Г. Системный анализ и моделирование опасных процессов в техносфере / Белов П.Г. – М.: Академия, 2003. – 512 с.
11. Астафьев А.Г. Техногенные аварии и катастрофы / Астафьев А.Г. – М.: Кнорус, 2006. – 255 с.
12. Четыркин Е.М. Статистические методы прогнозирования / Четыркин Е.М. – М.: Статистика, 1977. – 200 с.
13. Лапач С. Н. Прогнозирование с использованием полигональной регрессии / Чубенко А.В., Бабич П.Н. – Провизор. – 2003. – № 16.

О.А. Левтеров

Застосування генетичних алгоритмів при прогнозуванні факторів НС

Запропоновано використання генетичних алгоритмів при прогнозуванні факторів НС. Розглянуто методику розрахунку краткованого прогнозу розвитку площі обмеженою крайкою пожежі з використанням генетичного алгоритму.

Ключові слова: генетичний алгоритм, краткований прогноз, надзвичайна ситуація.

A.A. Levterov

Application of genetic algorithms when forecasting factors of a emergency

Use of genetic algorithms when forecasting factors of an emergency is offered. The method of calculation of the short-term forecast of the area, by a limited edge of the fire with use of genetic algorithm is considered.

Keywords: genetic algorithm, short-term forecast, emergency situation.